ACTIVATION OF A TRANSCRIPTION-DEPENDENT CRISPR-CAS SYSTEM

by
Kaitlin Johnson

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
June 2019

ABSTRACT

CRISPR-Cas provides prokaryotes with adaptive immunity from bacteriophage and other MGEs (mobile genetic elements). CRISPR-Cas systems use a crRNA-guided effector complex to specifically target and destroy invading nucleic acids. These immune systems are immensely diverse among the six types, and often within subtypes. Whereas most types of CRISPR-Cas target foreign DNA, Type III systems (Csm/Cmr) utilize a transcription-coupled mechanism to target RNA. Upon target RNA binding, the Type III effector degrades both RNA and DNA and produces a secondary messenger that activates a trans-acting RNase. To prevent inappropriate activation by host-derived targets, effector activity is dependent upon identification of the target's origin, which is signaled by the nucleotides most adjacent to the targeted sequence. Two mechanisms for this identification have been described in two subtypes of Type III systems. The complex may specifically recognize these sequences, or a host-derived sequence may base pair with the crRNA tag. It is currently unknown if the mechanism used by a given effector is subtype- or species-specific.

Contrary to data available for most other CRISPR-Cas systems, it has been suggested that Type III effectors are extremely tolerant of mismatches. The data supporting this is sparse, but mismatch tolerance has been demonstrated in one subtype of Type III. It is also unknown if this is a subtype- or species-specific characteristic.

Here we utilize recombinantly purified *Thermotoga maritima* Cmr (*Tma*Cmr) complex and RNA targets generated by *in vitro* transcription to systematically explore these questions. We demonstrate that both mechanisms are utilized by *Tma*Cmr to identify host-derived transcripts. We also show that *Tma*Cmr is extremely tolerant of mismatches. These results unify the features of the two major subtypes of Type III CRISPR-Cas systems.

ACKNOWLEDGEMENTS

I am so thankful for all of the people who have supported me during my PhD. Thank you to the BMB department for all of the amazing opportunities and friendships over the last six years. I am especially grateful to the following people for their support and encouragement. This body of work likely would not exist without them.

Thesis Committee

Jennifer Kavran, Jungsan Sohn, and Sean Prigge

Thank you for your advice during thesis committee meetings and for reading this thesis. Thank you all for being approachable, knowledgeable, and willing to help.

Bailey Lab

This is a wonderful lab. I am so grateful for being able to share reagents, protocols, and knowledge with all of the talented scientists in the Bailey lab, past and present.

Scott Bailey: Thank you for everything. You have been such a great mentor and the most fun PI to work for, and I can't imagine doing my PhD in any other lab. Thank you for your patience and tolerance. I truly appreciate how you have seemingly endless knowledge about science (and food, sports, television, etc.) and yet you are always approachable and have never made me feel inferior or dumb. I am grateful that you welcomed me into your lab even after I told you I wasn't interested in bacteria (and thank you for not laughing at me when I realized they're actually really cool). Thank you for being understanding and caring during my professional and personal struggles. You have taught me life skills that extend beyond the experimental and technical knowledge that I sought from this PhD program. I

know that your mentorship has shaped me as a scientist and as a person in ways for which I will be forever grateful and proud.

Brian Learn: You are BMB's fount of infinite wisdom. I have learned so much from working with you on this project. I am so grateful that you did all of those EMSAs; I don't know how long it would have taken me to optimize that assay, but I do know my gels would not have been as beautiful as yours.

John Mallon: I appreciate your willingness to help with anything, whether that may be making sense of data, imaging gels, fixing equipment, or making armor out of packing materials. Thank you for always being helpful, approachable, and selfless. You are the best labmate, a great teacher, and a dependable friend and I will miss working with you.

Anita Ramachandran: I am so grateful for your willingness to listen to me complain, for commiserating with me about our shared anxieties and frustrations, and most importantly for your support through some difficult times the last few years. Thank you for being my bay-mate, my therapist, and my friend.

Michael Estrella: Thank you for your guidance when I joined the lab, for your hard work in laying the foundation of this project, and for your meticulous lab notebook.

Friends and family

Grace Hwang: You may be the most loyal, caring, kind, and dependable person I have ever met and you always push me to be a better person. Thank you for always checking up on me, for supporting me, for not being afraid to call me out on my mistakes, and for always being willing to de-stress (usually by eating way too much food) with me.

Dad: I truly would not be who I am today without your support. Thank you for your advice and guidance.

Mom: Thank you for being my best friend, for teaching me what an independent woman looks like, and for having infinite patience with me. I admire you and appreciate everything you've done for me, the endless support you've given me, and the hours and hours you've spent on the phone with me.

Marc Kalma: You have done so much for me and I don't think I can ever repay you for that. Thank you for making me breakfast every day, for always making me laugh (often until I cry), for being patient and understanding when I'm grouchy, and for always being proud of me. It seems that all of my life's biggest challenges have come during the last three years, and I can't imagine how I would have managed any of them without you. Thank you for everything you do for me and for the wonderful person that you are.

Padme: Of course I have to acknowledge my cat for being the best companion. She is a constant source of entertainment and affection and always keeps me company during long nights of writing or making figures. Plus she's just so cute.

TABLE OF CONTENTS

LIST OF FIGURES

List of Abbreviations

ATP	adenosine triphosphate
Cas	CRISPR-associated
CARF	CRISPR-Cas associated Rossman fold
cOA	cyclic oligoadenylate
crRNA	CRISPR RNA
CRISPR	clustered regularly interspaced short palindromic repeat
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
MGE	mobile genetic element
PAM	protospacer adjacent motif
PFS	protospacer flanking site
Pfu	*Pyrococcus furiosus*
PMSF	phenylmethylsulfonyl fluoride
PNK	polynucleotide kinase
pre-crRNA	precursor crRNA
RNA	ribonucleic acid
rNTP	ribonucleoside triphosphate
rSAP	recombinant shrimp alkaline phosphatase
Sep	*Staphylococcus epidermidis*
ssDNA	single-stranded DNA
TCEP	tris(2-carboxyethyl)phosphine
Tma	*Thermotoga maritima*
tracrRNA	trans-activating crRNA
WT	wild-type

CHAPTER 1: CRISPR-CAS IMMUNITY

Viruses are ancient, ubiquitous, and diverse, surviving in all environments on Earth and infecting every known living organism[1]. Although not considered living, they are among the most successful biological entities, and are proposed to be the precursor to life[1]. Viruses that infect prokaryotes, known as bacteriophage (or simply phages), are incredibly numerous with an estimated population of 10^{31}, making bacteriophages the most abundant biological units on Earth[2,3]. Genetic diversity among bacteriophages is extraordinarily high, due to their long evolutionary history and the relative ease by which recombination and horizontal gene transfer can occur among such a vast population[2]. Prokaryotes and phages have coevolved over billions of years, each exerting selective pressure on the other and driving dynamic genetic change in their counterpart[3]. In a classic example of the Red Queen hypothesis, phages drive prokaryotes to evolve in order to avoid extinction, which in turn drives phages to counter-adapt to avoid their own demise[4,5].

Prokaryotic defense mechanisms

In order to survive a diverse bacteriophage population that outnumbers their own ten to one[5], prokaryotes have developed numerous forms of defense in order to protect themselves from viral invasion. The number of defense systems within prokaryotic genomes correlates more or less linearly with the size of the genome; the larger the genome, the more genes dedicated to protecting it[6]. Prokaryotic defense mechanisms fall into four main categories: invasion resistance, inhibition of phage replication, programmed dormancy or cell death, and immunity (Figure 1). The first line of bacterial defense, invasion resistance,

includes mechanisms by which bacteria prevent a virus from binding to the bacterial cell or from successfully injecting its genetic material; these strategies include reducing the expression of, altering, or blocking the bacterial receptor to which the phage adheres, as well as altering, blocking, or inhibiting the formation of the nucleic acid injection site[3,6]. Bacteria have also developed a wide array of mechanisms to prevent a bacteriophage from replicating, including blocking phage DNA integration or replication, inhibiting phage transcription, and preventing phage particle packaging or release[3,7].

If these and other immune strategies fail, an infected bacterium may become dormant or undergo programmed cell death in an attempt to confine the infection and protect the larger population. This is typically achieved through systems that inhibit essential cellular processes, either by inducing expression of a host protein that directly interferes with a metabolic process (abortive infection systems), or by inducing degradation of an antitoxin, which is required for repression of a toxin that interferes with one of a diverse group of cellular processes (toxin-antitoxin systems)[3,6,7]. These strategies are considered altruistic as they sacrifice the individual in order to preserve the population[7].

While dormancy or programmed cell death may help to prevent the spread of a bacteriophage among a population of bacteria, many prokaryotes have developed immune responses to defeat phage infections while avoiding cellular suicide. Immune systems involve destruction of invading nucleic acid and may be innate or adaptive[3,6]. Innate immunity, such as restriction-modification systems and likely argonaute-mediated systems, provides a nonspecific response to foreign genetic material. Restriction-modification systems encode a methyltransferase, which methylates host DNA, and an endonuclease (also called restriction enzyme), which cleaves unmethylated (foreign) DNA at sequence-specific

locations[3]. Previously thought to exist only in eukaryotes, recently discovered prokaryotic argonautes are proposed to serve as phage-defense systems by destroying foreign nucleic acid in a DNA or RNA guide-dependent manner; however, it is currently unclear how these guides are generated and no studies have yet confirmed this function[3,8]. Adaptive immune systems provide the host with a memory of prior infection in order to generate a specific response in the event of subsequent viral invasion. In the only known prokaryotic adaptive immune system, CRISPR-Cas, clustered regularly interspaced short palindromic repeat (CRISPR) arrays serve as a memory bank of previous viral encounters, providing RNA guides for the CRISPR-associated (Cas) nucleases[3].

CRISPR-Cas Immunity

A CRISPR locus contains Cas genes and a CRISPR array, which consists of the promoter-containing leader sequence, foreign-derived sequences (called spacers) interspersed with identical host sequences (called repeats)[9,10]. These repeating sequences were first reported in 1987[11] and spacer sequences were shown to be derived from mobile genetic elements (MGEs) in 2005, implicating a role for these systems in adaptive immunity[12–14]. This role was confirmed two years later when it was shown that new spacers derived from phage were incorporated into CRISPR arrays following phage challenge; with Cas genes, these spacers conferred resistance to subsequent phage infections[9]. This is accomplished through an RNA-guided effector complex; spacers are transcribed to form guide CRISPR-RNAs (crRNAs), which direct Cas nucleases to cognate nucleic acid sequences for destruction[15] (Figure 2).

CRISPR-Cas systems are found in about 50% of sequenced bacterial genomes and nearly 90% of archaea, and are nearly ubiquitous in hyperthermophiles[6,16]. There is

incredible diversity among CRISPR-Cas systems; although the ability to acquire new sequences makes CRISPR-Cas a relatively adaptable defense system, the constant evolution of phages creates the need for diversity among defense systems nonetheless[3,4]. As such, gene composition, effector complex organization, and mechanisms of interference vary widely among CRISPR-Cas systems. To date, two classes (1 and 2), six types (I-VI), and more than 20 subtypes of CRISPR-Cas interference systems have been identified[15]. Classes are defined by the organization of the interference complex; Class 1 complexes consist of a crRNA and multiple proteins, while Class 2 complexes contain a single guide RNA-bound protein[10,15]. Classification into the different types and subtypes is based on a number of factors, including the signature genes of each type/subtype, Cas protein sequence homology, locus architecture, and phylogeny of Cas1, the most conserved Cas protein[15]. Figure 3 summarizes the important features of each CRISPR-Cas type.

Cas1, found in all six types of CRISPR-Cas systems, is one of the key players in spacer acquisition, the first step of CRISPR-Cas immunity[15,17] (Figure 2). The acquisition complex is typically formed by a heterohexameric complex of four catalytic Cas1 subunits and two Cas2 subunits, which perform a structural function[10,17]. This acquisition machinery is responsible for capturing foreign DNA (protospacers) and integrating them as new spacers in the CRISPR array. Cas1-Cas2 binds fragments of viral DNA that contain a specific sequence called a protospacer adjacent motif (PAM), which does not exist within the CRISPR array[17]. These fragments are likely produced by DNA repair machinery; the free dsDNA end (which typically signal a double strand break in circular bacterial genomes) of the viral DNA recruits RecBCD, which resects the ends, producing short pieces of DNA[17]. Upon protospacer capture, the Cas1-Cas2 complex inserts the protospacer adjacent to the

first repeat by catalyzing two cleavage-ligation reactions[17]. This places the new spacer at the leader end of the array, thus imparting a chronological history of infection; this placement (nearest the promoter) likely provides more robust generation of the crRNA that targets the most recent invader, which tends to be the cell's biggest threat[17,18]. There are two modes of acquisition: naïve acquisition, in which new spacers are incorporated, and primed acquisition (or simply priming), which incorporates spacers that share homology with existing spacers[17]. Priming provides an enhanced rate of acquisition (relative to naïve acquisition)[17]. This reduces viral escape in two ways; priming stimulates integration of targets in the event of viral mutation and allows for insertion of a copy of an existing spacer nearer to the promoter to allow for more robust crRNA generation[17,19].

The second step of CRISPR-Cas immunity is crRNA generation (Figure 2). The CRISPR array typically contains one transcription start site within the leader sequence, thus producing a single precursor crRNA (pre-crRNA). The pre-crRNA is processed to generate multiple mature crRNAs through endonucleolytic cleavage within the repeat sequences[10]. The nucleases involved in crRNA maturation vary; in some types of CRISPR systems, the processing nuclease is part of the interference complex, while in others this function is performed by a trans-acting Cas or non-Cas nuclease[10,15]. Mature crRNAs consist of a small segment of repeat sequence on the 5' end, referred to as the 5' tag (or crRNA tag), followed by the spacer sequence. Cas proteins identify the tag to assemble around the crRNA, and the spacer within the crRNA guides the complex to its targets to complete the final step of immunity: interference (Figure 2).

CRISPR complexes may be guided by their crRNA to cognate DNA, as in Types I, II, V, and likely IV, or cognate RNA, as in Types III and VI[15]. Some interference complexes

are also capable of degrading collateral nucleic acids upon binding to targets; a nonspecific DNA nuclease domain is activated in Type III complexes upon target binding, while target binding in Type VI complexes activates a second, nonspecific RNA nuclease domain[15,20,21]. To avoid inappropriate targeting, a proper target must be complementary to the crRNA and must be identified as foreign in origin in order to activate the interference complex. A crRNA is naturally complementary to the DNA from which it is derived; thus DNA-targeting complexes must confirm the target is foreign DNA to avoid destruction of the host genome (autoimmunity). To this end, complexes in Types I, II, and V identify foreign sequences by a specific protospacer adjacent motif (PAM), which is a short sequence that is not found in the CRISPR array[10]. Similarly, RNA-targeting complexes must avoid activation by anti-CRISPR transcripts, which can arise from the anti-sense transcription of a CRISPR array[22]. In Types III and VI, complementarity between the tag of the crRNA and the protospacer flanking site (PFS) is a signal of host transcript and prevents interference complex activation[20,23]. Host target binding in Type VI prevents both targeted and collateral RNA cleavage[20], while in Type III, host target RNA is cleaved and only collateral DNA cleavage is inhibited[23].

Class 1 CRISPR-Cas interference

While performing similar functions in a crRNA-guided manner, CRISPR-Cas interference complexes themselves are quite diverse. Class 1 effector complexes are composed of multiple proteins, typically four to seven subunits in an uneven stoichiometry, bound to the crRNA guide[24]. Class 1 systems are considered the evolutionary ancestors of CRISPR-Cas; genomics studies indicate that the ancestral Cas effector likely utilized a

progenitor of the Type III Cas10 subunit[24]. Class 1 systems are widespread in both archaea and bacteria.

Type I CRISPR-Cas interference

Type I CRISPR-Cas systems are the most common, accounting for about 60% of sequenced CRISPR-Cas systems in both bacteria and archaea[16]. With at least seven subtypes, they are also most diverse[6]. The subunits of the Type I complex, Cascade (CRISPR-associated complex for antiviral defense), are so diverse among the subtypes that subunit homology is often established based on function, rather than sequence[10,16]. However, the overall architecture and function of the Type I interference machinery is conserved. The pre-crRNA forms stem-loop structures with the repeat, which are recognized by Cas6 for crRNA maturation[10]. After processing the crRNA, Cas6 remains bound to the hairpin on the 3' end, while Cas5 and Cas8 (the large subunit) cap the 5' repeat[10,25]. Multiple Cas7 subunits filament along the crRNA, each with a palm domain to stabilizing five nucleotides of the crRNA backbone and a thumb domain that kinks the crRNA at every sixth nucleotide; this allows for efficient base pairing between the crRNA and target[25]. Two copies of Cas11 (the small subunit) interact with the Cas7 subunits to stabilize the nontarget strand[10]. There are many subtype-specific differences in the Type I interference machinery (for example, in several subtypes, the small subunit is fused to or functionally replaced by the large subunit), but the overall structure of each tends to resemble a seahorse, with Cas6 forming the head, Cas7 acting as the backbone, Cas11 creating the belly, and Cas5 and Cas8 making up the tail[10,25].

Cascade scans double-stranded DNA for PAM sequences; the large subunit recognizes PAM and melts a small region of the duplex[10]. This local unwinding exposes a small region of single-stranded DNA near the identified PAM, which allows the crRNA to sample for complementarity. If the DNA is not complementary to the first eight nucleotides (called the seed sequence) of the crRNA, Cascade dissociates and continues scanning for PAM. Upon identifying a target with complementarity to the crRNA seed, Cascade directionally unwinds the DNA. The target strand forms a duplex with the crRNA and the nontarget strand is held by the small subunit to stabilize the subsequent R-loop[10,26]. Cascade recruits a trans-acting helicase-nuclease, Cas3, which nicks the displaced nontarget strand[26]. Cas3 then undergoes a conformational change that activates the helicase domain and translocates and degrades 200-300 nucleotides of DNA[10,26]. Another Cas or non-Cas nuclease may further degrade the target genome[10]. A schematic summary of Type I interference can be found in Figure 4A.

Type III CRISPR-Cas interference

Type III CRISPR-Cas systems account for 34% and 25% of sequenced CRISPR-Cas systems in archaea and bacteria, respectively[16]. There are four characterized subtypes[15]. Although not as prevalent as Type I systems, with three separate nuclease functions, it is the most complex of the CRISPR-Cas types. While functionally distinct, Type III effector complexes, called Csm or Cmr depending on the subtype, are structurally similar to Cascade. Like in Type I systems, pre-crRNA is processed by Cas6, which identifies the repeat sequence and cleaves eight nucleotides upstream of the spacer, leaving each crRNA with a segment of repeat (the crRNA tag) on the 5' end[27]. However, Cas6 dissociates from mature

crRNA and is not a part of Type III effector complexes[21]. In most subtypes, further processing by host nucleases removes repeat sequence from the 3' end of the crRNA, leaving a mature crRNA with just a 5' tag and a spacer[28]. Csm4/Cmr3 and the large subunit, Cas10, surround the tag[10]. Csm3/Cmr4 form the backbone and filament on the crRNA spacer, with multiple copies of the small subunit, Csm2/Cmr5, at the belly. Like in Cascade, the thumb domains of the backbone subunits protrude into the crRNA spacer, flipping out every sixth nucleotide[29]. The number of backbone and small subunits in the complex depends on the length of the crRNA[28]. The 3' end of the crRNA is capped either by Csm5 or by Cmr1 and Cmr6[21,28].

The Csm/Cmr complex binds to RNA targets that are complementary to the crRNA within the complex. Data suggest that target RNA:crRNA duplex formation does not rely upon a seed region[30]. Foreign target binding induces a conformational change and initiates three enzymatic activities within the complex simultaneously[21,28,31–33] (Figure 4B). (1) The Csm3/Cmr4 subunits cleave the bound RNA target at six nucleotide intervals[21,28]. (2) The HD domain of the Cas10 subunit is activated to nonspecifically degrade adjacent single-stranded DNA by cleaving after thymine residues[21,28]. (3) The Palm domain of Cas10 is activated to convert ATP to cyclic oligoadenylate (cOA), which acts as a secondary messenger[31,32]. A fourth enzymatic function, this one outside of the complex, is then activated by cOA production; upon binding to cOA, Csm6/Csx1 is activated to nonspecifically degrade RNA[31,32]. The binding site for cOA is within the CARF domain (CRISPR-Cas associated Rossman fold) of Csm6/Csx1[31,32]. The functions of other CARF domain proteins, many of which are associated with Type III CRISPR-Cas systems (and

some with Type I), are unknown[34], but it is likely that these could be regulated by cOA as well.

The proposed model is that Type III CRISPR-Cas interference occurs during transcription[21,28]. The Csm/Cmr complex may bind to a complementary transcript as it is being produced by RNA polymerase; this would provide access to both the RNA target and foreign ssDNA within the transcription bubble[21,28]. The complex is temporally regulated by RNA target degradation and dissociation; thus when the transcript is cleared from the complex, the structure reverts to its original conformation, and cOA production and DNA cleavage cease[22,35,36]. This minimizes the likelihood of off-target DNA cleavage.

The ability to identify host-derived transcripts also helps Csm/Cmr to avoid off-target DNA cleavage. RNA cleavage by Csm/Cmr requires complementarity between the crRNA spacer and target protospacer, and proceeds regardless of transcript origin[36]. However, identification of self RNAs locks the Cas10 in an inactive state, preventing activation of the HD and Palm domains (and therefore cOA production and DNA cleavage)[33]. While it is unknown how prevalent host-derived transcripts might be, they could conceivably arise from the antisense transcription of the CRISPR array, particularly if a viral promoter was incorporated as a spacer[22]. While, like a target RNA, the protospacer of these transcripts would be complementary to the crRNA spacer, the 3' protospacer flanking site (PFS) would also be complementary to the crRNA tag; this complementary PFS is called an anti-tag. Foreign transcripts do not contain anti-tag sequences, which makes this a good method by which to distinguish host from foreign RNA. This discrimination may occur by two mechanisms: identification of the anti-tag sequence[37] or base pairing between the anti-tag and tag sequences[30,38].

Type IV CRISPR-Cas interference

There is little known about Type IV CRISPR-Cas systems. This single subtype system is rare, making up less than 2% of CRISPR-Cas loci[16]. Type IV loci are often found on plasmids and are often not adjacent to a CRISPR array[16]. These loci also lack adaptation machinery[15]. It is thought that Type IV systems target DNA[15]; however thus far, no experimental interference data are available. A recent study on the assembly of a Type IV complex provides insight into crRNA processing and the structure of the effector (Figure 4C). A Cas6-like nuclease, Csf5, cleaves the pre-crRNA, leaving a repeat tag at the 5' end of the crRNA and a hairpin on the 3' end[39]. Csf5 remains bound to the 3' end of the crRNA and Csf2 forms the backbone of the complex[39]. Csf1 is the large subunit homolog, and Csf3 forms the 5' cap on the crRNA[39].

Class 2 CRISPR-Cas interference

Class 2 CRISPR-Cas systems account for less than 15% of bacterial CRISPR loci and are rarely found in archaea[16]. Containing a single bilobed protein, Class 2 effectors are simpler than Class 1 complexes[15]. Due to their simplicity, these crRNA-directed nucleases have garnered much interest for their utility in molecular biology and biotechnology[40].

Type II CRISPR-Cas interference

The three subtypes of Type II CRISPR-Cas systems together make up about 13% of bacterial CRISPR loci[16]. Type II has been found in only one archaeon[15,16]. Although not found in a great many organisms (5% of all bacteria utilize this immune system[41]), the simplicity of Type II systems and their consequent utility as a programmable biological tool has made them the subject of extensive studies. These effectors are made of a single protein,

Cas9, and two RNA guides[42]. Like all CRISPR-Cas systems, Cas9 requires crRNA to identify complementary targets, but a trans-activating crRNA (tracrRNA) is also required for crRNA maturation and target cleavage. The tracrRNA is encoded within or near the CRISPR array and is complementary to the crRNA repeat[43]. Prior to processing the pre-crRNA and tracrRNA form a duplex, which is recognized and bound by Cas9; this leads to the recruitment of a bacterial nuclease, RNase III, which cleaves within the RNA duplex, and a second unknown nuclease that removes the 5' segment of repeat to complete crRNA maturation[10].

Assembly of Class 1 effectors requires the presence of crRNA around which the individual subunits scaffold, but as Class 1 effectors are single proteins and do not require such assembly, they must be held in a catalytically inactive conformation until the guide RNA duplex is bound[15,42]. Loading of tracrRNA:crRNA induces a dramatic conformation change that creates a channel to accommodate target DNA and rearranges the C-terminal domain (CTD), which is disordered in the unbound state[26,42]. The guide RNA resides within a positively-charged groove between the two lobes of Cas9: the recognition (REC) lobe and the nuclease (NUC) lobe[24]. Unlike Class 1 effectors, which preorder the entirety of the crRNA for duplex formation, Cas9 only preorders the 10-12 nucleotides in the seed region, likely due to topological constraints (which are relieved in Class 1 complexes by the flipped out nucleotides and helical structure)[42]. The alpha-helical REC lobe interacts with the crRNA:target DNA duplex, while the NUC lobe contains two nuclease domains, RuvC and HNH[26].

Much like Cascade, Cas9 scans dsDNA for PAM; sequences are interrogated by the CTD and recognition leads to unwinding of the dsDNA adjacent to the PAM[42]. The

preordered seed region can then probe the PAM-proximal nucleotides, and complementarity within the seed prompts a conformation changes within Cas9 to allow full duplex formation with the target and to stabilize the resulting R-loop[10]. Further complementarity induces yet another conformation change that activates the HNH nuclease domain to cleave the target strand and allosterically activates the RuvC nuclease domain to cleave the nontarget strand[26]. These concerted cleavage events lead to a double-strand break located three base pairs upstream of the PAM sequence[10] (Figure 5A).

Type V CRISPR-Cas interference

Type V immunity is rare, representing less than 2% of CRISPR-Cas systems[16,41]. Nevertheless, Type V systems are quite diverse and ten subtypes have been proposed[44]. The effector proteins from these subtypes, Cas12, are quite diverse but have conserved catalytic motifs associated with a RuvC-like endonuclease domain; due to this domain and some functional similarities, Type V systems were originally classified as a subtype of Type II CRISPR-Cas[15,16]. However outside of the shared nuclease motifs, Cas12 and Cas9 proteins share no sequence or structural homology and it is now thought that Type II and Type V evolved separately to converge upon a similar function[15]. In fact, because of the diversity within the RuvC-like domains (as well as throughout the rest of the proteins), it is likely that many subtypes of Type V also evolved independently from multiple transposon-encoded nucleases[10,44]. Support for this theory was recently demonstrated in a study that included both metagenomic analysis of CRISPR-Cas systems and biochemical analysis of several functionally diverse Type V systems[44]. Most Cas12 effectors are thought to target dsDNA, generating a double-strand break and further degradation of each strand, but this study demonstrated RNA targeting with collateral RNA and ssDNA degradation in one particular

subtype, as well as dsDNA nicking in another[44]. The model described below is based largely on Cas12a (of Type V-A), the most well studied Type V effector (see also Figure 5B).

About half of Type V subtypes utilize a tracrRNA[44]. Dual nuclease activity allows Cas12 (with tracrRNA where required) to process pre-crRNA[10,44]. Like Cas9, Type V effector proteins are inactive until bound to their guide RNAs[26]. A conformation change induced by guide RNA binding allows Cas12 to search for PAM, and upon recognition, locally unwind the dsDNA to expose the first five nucleotides and test for seed complementarity[45]. The rest of the R-loop is then formed and a conformational change leads to the opening of a catalytic pocket[45]. A single active site creates two successive nicks, one on each strand of the DNA target, generating a staggered break in the dsDNA and leaving short 5' overhangs[15,44,46]. Both the target and nontarget strand can then be degraded as the conformational change induced upon target binding enables nonspecific ssDNA shredding by the same domain[46]. The conformation of Cas12 does not revert to the pre-activated state as long as the R-loop remains bound, leaving the catalytic pocket open to complete destruction of the bound target or any adjacent ssDNA[45,46]. The nontarget strand is cut much more efficiently than the target strand due to its positioning within the catalytic pocket, which allows the R-loop to persist for extensive ssDNA degradation[45]. The effector conformation is reset upon crRNA displacement; a new crRNA molecule replaces the crRNA:target duplex, reverting Cas12 to its inactive state[45].

Type VI CRISPR-Cas interference

Type VI is the most recently identified CRISPR-Cas system[47]. It represents less than 1% of CRISPR-Cas loci in sequenced bacterial genomes and has not been found in archaea[41]. Four subtypes of Type VI systems have been proposed[15,48]. These exhibit great

diversity; the effector, Cas13, and in particular the two HEPN domains that define Type VI systems, show even greater sequence divergence than the other Class 2 systems[15]. Like Cas12, Cas13 also processes the pre-crRNA; however none of the Type VI subtypes require a tracrRNA[10,15]. The crRNA is bound within a cleft of the bilobed structure that is common to Class 1 effectors[48]; upon its binding, Cas13 undergoes a conformational change to permit duplex formation between the crRNA and target RNA[10]. Cas13 does not require a PAM sequence, and instead relies on the PFS for host target identification[10,20]. Additionally, Cas13 does not utilize a canonical seed sequence, but does require a minimum complementarity between crRNA and RNA target[48].

Target identification induces a conformational change that brings the active site residues of the two HEPN domains into proximity with one another, forming a bipartite active site that faces the exterior of the protein[48]. This solvent-exposed RNase active site readily degrades the bound target and other adjacent RNAs[10,48] (Figure 5C). The indiscriminate RNA cleavage has been shown to degrade host RNAs and restrict bacterial growth in heterologous systems; this indicates that Type VI effectors may induce dormancy or cell death or that the natural host employs other strategies to keep this RNA degradation in check[10].

Unlike Type III systems which degrade self RNA targets but inhibit the collateral DNase activity in their presence, complex formation between the Cas13 crRNA and a self RNA target leads to inhibition of both specific and nonspecific RNA cleavage[20]. The identification of self RNA occurs at the PFS, with some systems requiring specific sequences in the 3' PFS (adjacent to the crRNA tag) and some seeming to have requirements in both 5' and 3' PFS[10]. It has been demonstrated in one subtype that the mechanism of PFS

identification is through tag:antitag complementarity and proposed that this could be a universal mechanism of self sequence identification by Type VI effectors[20]. Some Type VI effectors do not require processing of their crRNA, which would leave repeat sequence on both sides of the spacer[10]; this may explain why some subtypes have a sequence requirement in both PFSs.

Figures

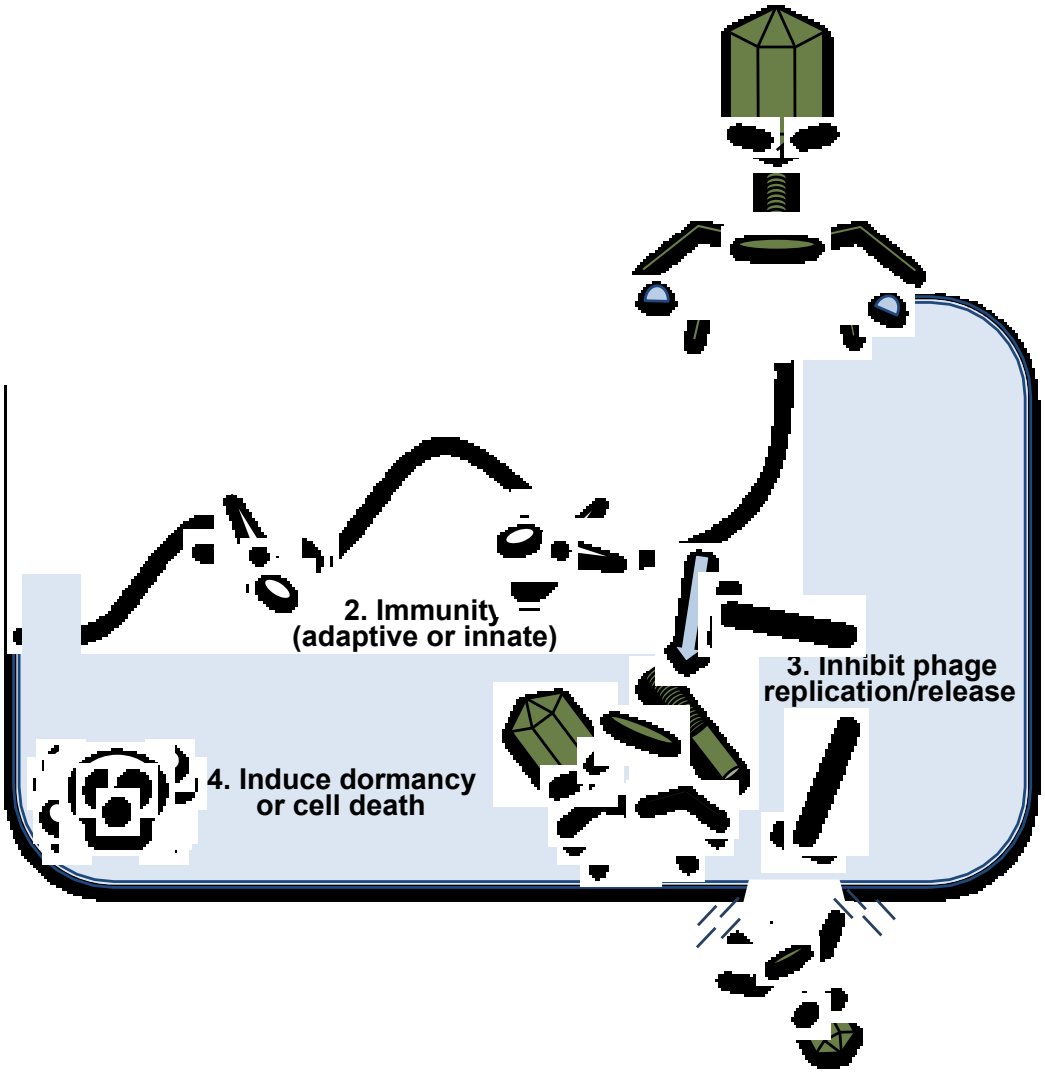

Figure 1. Mechanisms of phage resistance in prokaryotes. (1) By preventing a phage from adhering to the cell surface or injecting its genetic material, invasion is resisted. (2) The ability to destroy nucleic acid provides adaptive or innate immunity. (3) Preventing viral transcription, translation, or packaging, inhibits viral replication. (4) To prevent spread of the phage and protect surrounding cells, the cell may sacrifice itself by inducing cell death or dormancy.

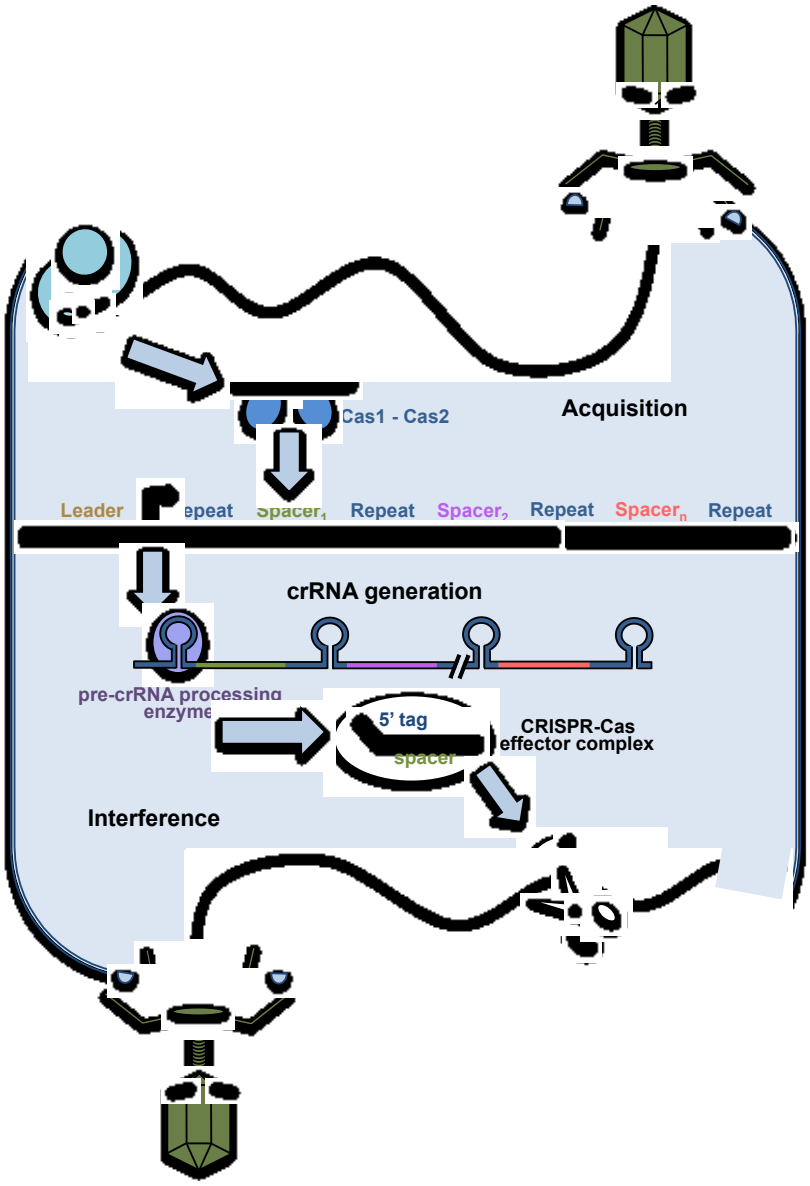

Figure 2. General mechanism of CRISPR-Cas immunity. *Acquisition*: *the process by which new spacers are incorporated into the CRISPR array.* (1) Fragments of viral DNA are produced, likely via end resection by the host RecBCD. These fragments are captured by the acquisition machinery, Cas1-Cas2. (2) The fragment is inserted into the CRISPR array and becomes the leader-proximal spacer to provide immunity upon repeat phage exposure. *crRNA generation: the process by which the crRNA guides are produced* (3) Transcription is initiated at the transcription start site within the leader sequence and a single RNA, the pre-crRNA, is produced. (4) The crRNA is processed by Cas or other host nucleases to produce a mature crRNA around which the CRISPR-Cas effector complex forms. *Interference: the targeting and destruction of specified viral sequences.* (5) The crRNA guides the effector complex to viral targets. Activation of the effector complex requires verification that the target is of foreign origin (e.g. by identifying the PAM sequence) and complementarity between the crRNA and the target protospacer. Upon activation, the effector cleaves the target. See figures 3 and 4 for interference mechanisms employed by the different types of CRISPR-Cas effectors.

	CLASS 1			CLASS 2		
	Type I	Type III	Type IV	Type II	Type V	Type VI
Subtypes	I-A, I-B, I-C, I-D, I-E, I-F, I-U	III-A, III-B, III-C, III-D	IV	II-A, II-B, II-C	V-A, V-B, V-C, V-D, V-E	IV
Effector complex	 Cascade	Csm/Cmr	Csf	Cas9	Cas12	Cas13
Nuclease	Cas3	Cas10 Csm3/Cmr4	*Csf1*	Cas9 RUVC + HNH domains	Cas12 RUVC domain	Cas13 2 HEPN domains
crRNA processing	Cas6	Cas6	Csf5	RNase III	Cas12	Cas13
Target	DNA	DNA RNA*	*DNA*	DNA	DNA DNA*	RNA RNA*
Seed region	Yes	No	*?*	Yes	Yes	No
Self vs nonself	PAM	PFS	*?*	PAM	PAM	PFS

Figure 3. Defining characteristics and features of the six CRISPR-Cas types. *Italics* indicate proposed but not yet confirmed. Asterisk (*) indicates collateral targeting.

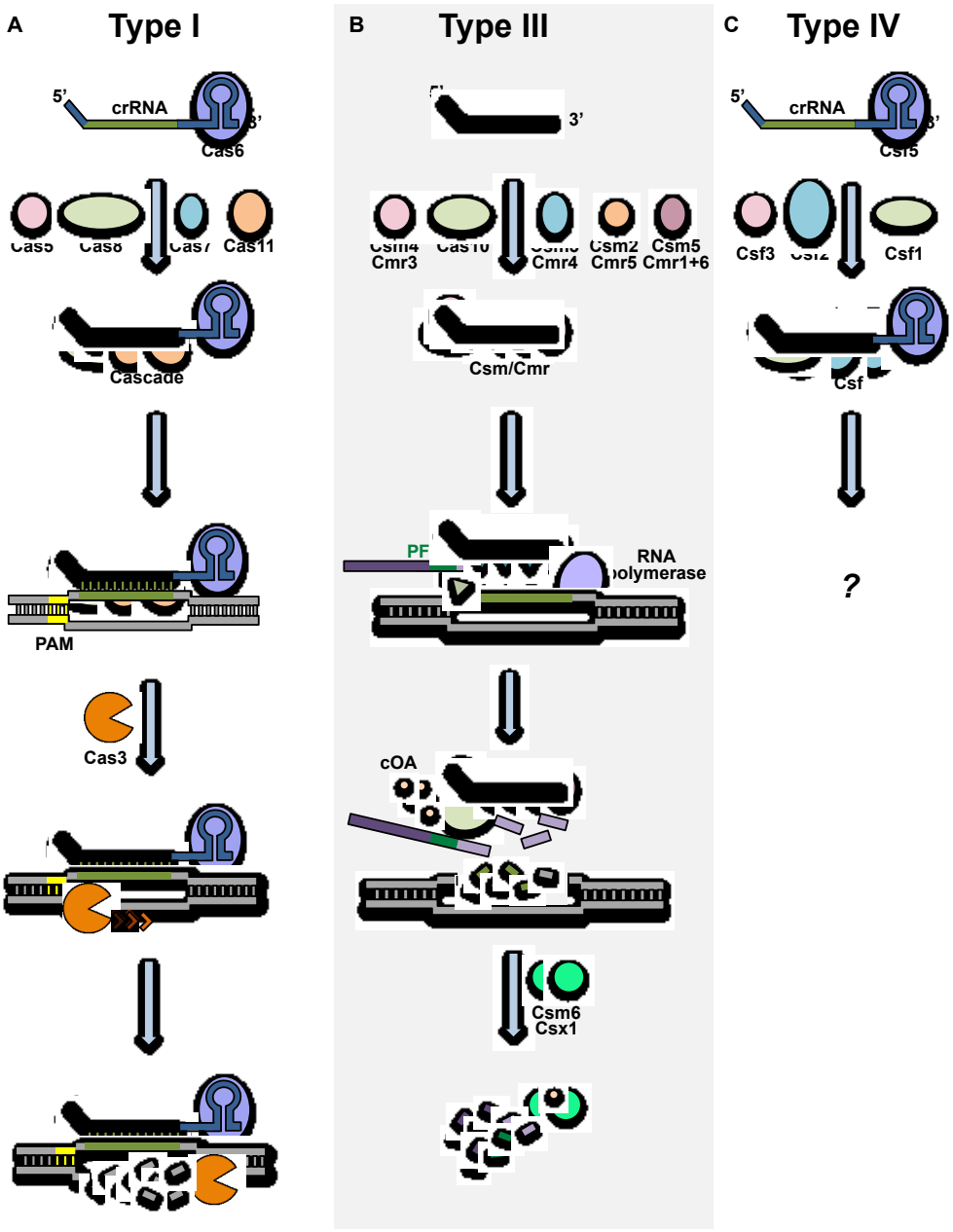

Figure 4. Class 1 interference. (A) Type II. Upon PAM recognition, Cascade melts the DNA to identify targets complementary to the crRNA. Target identification leads to the recruitment of Cas3, a trans-acting helicase/nuclease, which degrades the ssDNA. (B) Type III. The Csm/Cmr complex binds to and determines the origin (via the PFS) of complementary RNA targets. Foreign target binding leads to the activation of the three enzymatic functions of Csm/Cmr: (*1*) RNA is cleaved by the Csm3/Cmr4 subunits (*2*) adjacent ssDNA is cleaved by Cas10 in a sequence nonspecific manner, and (*3*) cOA is produced by Cas 10. A trans-acting RNase, Csm6/Csx1, is activated upon cOA binding and nonspecifically degrades RNA. (C) Type IV (proposed structure). The stoichiometry of Csf2 is yet unknown. Interference has not yet been demonstrated in Type IV and thus is not represented here.

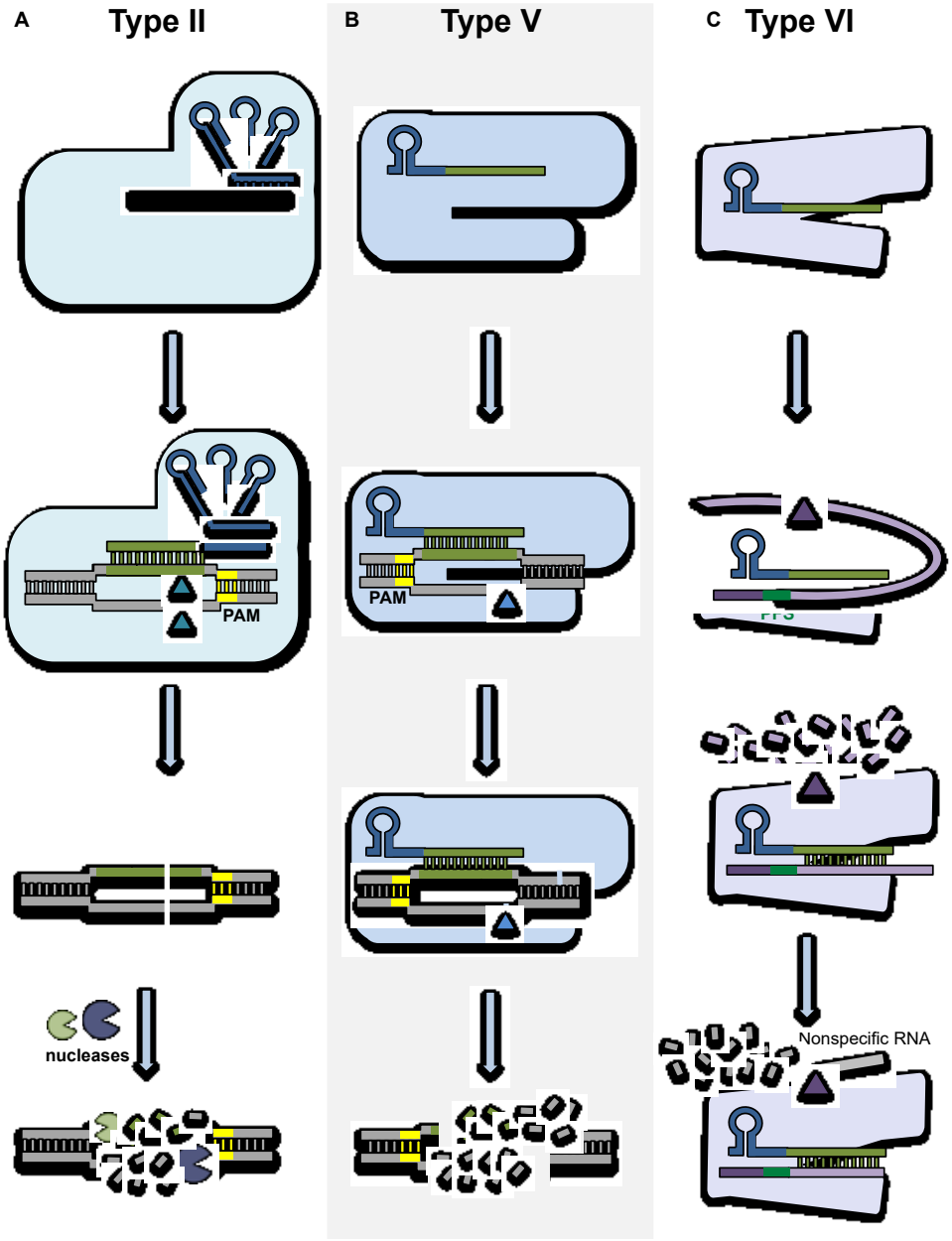

Figure 5. Class 2 interference. (A) Type II. Both crRNA processing and interference require tracrRNA, which consists of several hairpins and a region of complementarity to the crRNA tag; these two together form the RNA guide for the Type II effector, Cas9. Cas9 searches for PAM sequences and then melts adjacent dsDNA. Complementarity between the crRNA and target prompts the two nuclease domains of Cas9 to nick each strand of the DNA, creating a double-strand break. This break is then degraded by host nucleases. (B) Type V. This type exhibits extreme functional diversity, which is just now being uncovered; thus one subtype is represented here (Cas12a). The complex searches for PAM sequences and melts adjacent dsDNA. Complementarity between the crRNA spacer and target protospacer activates the nuclease domain to cleave both target and nontarget DNA strands, generating a staggered break. This nuclease degrades both strands, as well as any adjacent ssDNA.(C) Type VI. The complex targets complementary RNA, which Cas13 identifies as a foreign transcript by its PFS. This activates the RNase domain, which degrades both the target RNA and adjacent nonspecific RNAs.

CHAPTER 2: FOREIGN TARGET IDENTIFICATION IN A TYPE III-B CRISPR-CAS SYSTEM

Type III CRISPR-Cas systems offer robust protection from phage by utilizing a CRISPR-RNA (crRNA)-guided effector complex to degrade foreign RNA and DNA. The ability of these complexes to correctly identify RNA is essential for appropriate activation of the DNA degradation domain; this regulation is key to avoiding both autoimmunity (degradation of the host genome due to an incorrect signal) and immune suppression (successful invasion of phage due to deficient degradation). Two different mechanisms of self RNA identification have been described in Type III CRISPR-Cas systems: multiple type III-A studies have shown that self RNAs are recognized by base pairing between the target antitag and crRNA tag sequences, while a type III-B study indicated sequence-specific recognition of the RNA PFS. Here, we present evidence of both mechanisms playing a role in RNA identification in the type III-B (Cmr) system of *Thermotoga maritima*. Utilizing recombinantly purified Cmr complex and RNA targets generated through *in vitro* transcription, we systematically studied the role of antitag sequence in the target PFS. We find that, like the *Pfu*Cmr complex, *Tma*Cmr specifically recognizes a portion of the PFS; a guanosine in the first position of the PFS is an important indicator of a host-derived RNA. We also show that, like several Type III-A complexes, *Tma*Cmr recognizes base pairing between the PFS and the crRNA tag (at positions -2 to -5) to identify self transcripts.

Introduction

CRISPR-Cas provides prokaryotes with adaptive immunity from bacteriophage and other MGEs (mobile genetic elements)[9]. In competition with the rapid evolution of phage, CRISPR-Cas systems have diversified immensely among the six types[24]. These acquired immune systems are found in about 50% and 90% of bacteria and archaea respectively, and are nearly ubiquitous in hyperthermophiles[6,16]. Of these, about 25% and 34% respectively are Type III systems[16]. This type includes four subtypes: Types III-A and III-D which utilize the Csm effector complex, and Types III-B and III-C, which utilize the Cmr complex[16]. Whereas most types of CRISPR-Cas target foreign DNA, Type III systems (Csm/Cmr) utilize a transcription-coupled mechanism to degrade both RNA and DNA[10].

By coupling targeting to transcription and by degrading both forms of nucleic acid, Type III systems provide an extremely robust immune response and yet tolerate potentially beneficial MGEs. MGEs may contribute genes that provide a fitness advantage to the bacterium, such as genes that inhibit other phage invasions or increase bacterial pathogenicity and survival[49]. Incorporation of spacers from viral genes that are harmful to the bacterium, such as those that promote viral replication or bacterial lysis, allows for coexistence of bacterium and MGE until it becomes detrimental for bacterial survival[50]. Furthermore, as RNA complementary to the crRNA activates Type III effectors (and therefore the CRISPR array in the genome is not targeted), specific PAM sequences are not required for foreign DNA identification. Instead, Type III systems rely on the identity of the target RNA PFS for recognition of self RNA to prevent inappropriate activation of the interference machinery; anti-tag sequence in the crRNA-adjacent PFS prohibits activation of the Type III immune response[23]. This allows for broad targeting relative to that afforded by

Types I and II, where a mutation in the PAM sequence of the target allows the phage to escape targeting[10].

Like all CRISPR-Cas systems, the leader sequence upstream of the CRISPR array directs transcription of pre-crRNA as one RNA containing multiple spacers; in Type III systems, this pre-crRNA is processed by Cas6[21]. While in other CRISPR-Cas types the repeat sequence readily forms stem-loops that are identified by Cas6, many Type III repeats are predicted to form weak secondary structures, if any[51]. It is thought that in these cases Cas6 is able to stabilize a hairpin that is otherwise too unstable to form[51]. Cleavage by Cas6 at the base of the stem-loop leaves an eight-nucleotide segment of repeat sequence (the crRNA tag) at the 5' end of each spacer[27]. Further processing by other unknown nucleases may cleave the repeat sequence (or a portion of it) from the 3' end of the spacer[28]. Other Cas proteins, such as Csm2 and Csm5 in Type III-A, are often required for crRNA maturation[52].

The Csm/Cmr complex forms around the crRNA, with the length of the crRNA dictating the number of filamenting subunits (Csm3/Cmr4)[21,28]. Complexes often assemble around crRNAs that vary in length, with one added molecule each of Csm3/Cmr4 and Csm2/Cmr5 for every six additional nucleotides; however the function of these differently sized crRNAs and complexes is unknown[21,52,53]. Upon identification of an RNA target that is complementary to the crRNA spacer, the RNase domain in each Csm3/Cmr4 subunit cleaves the RNA, producing a six-nucleotide cut pattern[36,54–56]. A foreign RNA target also induces a conformational change that allows for activation of two domains of the Cas10 subunit[33,53,57]; the HD domain is activated to degrade ssDNA[22,36,37], and the Palm domain initiates production of a secondary messenger (cyclic oligoadenylate, or cOA) by the Cas10 Palm domain[31,32]. Figure 4B provides a schematic summary of Type III interference.

It was originally thought that Csm complexes targeted DNA[38,58], while Cmr complexes targeted RNA[54–56]. Once thought to be a difference among subtypes, this discrepancy likely arose due to a difference in experimental methods; initial studies performed with Csm complexes were done *in vivo*, while those with Cmr complexes were done *in vitro*. Both Csm and Cmr have since been shown to degrade both RNA and DNA, resulting in a unified model of transcription-coupled targeting[22,36,37].

In both Csm and Cmr, DNA cleavage activation is inhibited upon identification of self RNA targets; in both cases, an RNA target containing an antitag PFS does not allow Cas10 activation[36–38,59–61]. However the method by which this inhibition occurs may differ between Csm and Cmr. Type III-A (Csm) studies from *Staphylococcus epidermidis* (*Sep*) have indicated that base pairing between the crRNA tag and the target antitag prevents Cas10 activation[33,38]. Structural studies indicate that base pairing is possible between the antitag and tag sequences at positions -2 and -5; tag positions -1 and -6 to -8 are unavailable for base pairing and are therefore not likely to be involved in RNA identification in these systems[53,57].

By contrast, a study of a Type III-B (Cmr) system from *Pyrococcus furiosus* (*Pfu*) suggests that the first three nucleotides of the PFS are recognized specifically to allow or prevent Cas10 activation[37]. It is currently unclear if the mechanism of discrimination is subtype-specific (i.e. that Type III-A/Csm systems use base pairing, while Type III-B/Cmr systems use sequence specificity) or if either mechanism can be utilized by any given Type III system.

In this study, we investigate the mechanism of self and nonself target discrimination by the Cmr complex from *Thermotoga maritima* (*Tma*). Utilizing recombinantly purified *Tma*Cmr complex, we tested targets containing varying positions of antitag sequence within

the PFS to understand the mechanism of target discrimination. We determined that antitag sequence in positions -1 to -3 of the PFS is minimally sufficient to inhibit DNA cleavage by *Tma*Cmr. Alternatively, antitag sequence in positions -2 to -5 is also sufficient for inhibition of DNA cleavage. We find that, like the *Pfu*Cmr complex, *Tma*Cmr specifically recognizes a portion of the PFS; a guanosine in the first position of the PFS is an important indicator of a host-derived RNA. We also show that, like several Type III-A complexes, *Tma*Cmr recognizes base pairing between the PFS and the crRNA tag (at positions -2 to -5) to identify self transcripts.

Results

Activation of DNA cleavage requires a noncomplementary 3' protospacer flanking sequence

Activation of DNA cleavage by Type III systems requires pairing of the crRNA with an RNA protospacer and is regulated by the PFS. DNA cleavage is blocked in all Type III systems if the RNA target contains an anti-tag sequence in the PFS[22,36–38,59–61] (Figure 7A). However, the precise role of the PFS in the activation of DNA cleavage is unclear. Our lab previously reported that RNA targets lacking flanking sequences can activate DNA cleavage by *Tma*Cmr[36], but in other Type III systems, a PFS that lacks antitag sequence is required for activation[59,60]. To investigate this further, we monitored cleavage of ssDNA by *Tma*Cmr in the presence of a series of RNA targets. Our lab also previously showed that *Tma*Cmr cleaves ssDNA after thymine bases[36]; thus for simplicity of readout, we used a ssDNA oligonucleotide with a single thymine at its center. The *Tma*Cmr:crRNA complex was assembled from recombinant proteins (Figure 6A) and crRNA generated by *in vitro* transcription and subsequent processing with recombinant *Tma*Cas6 (Figure 6B-C). We

incubated labeled ssDNA substrate with the *Tma*Cmr:crRNA complex and excess RNA target composed of a complementary protospacer and 10 nucleotides of noncomplementary sequence at each flank. DNA cleavage was monitored by denaturing polyacrylamide gel electrophoresis (PAGE) followed by autoradiography. The expected cleavage product was observed after one minute at 80°C when all the required components (crRNA, *Tma*Cmr, complementary RNA target, and Mn^{2+}) were present (Figure 7B).

The fraction of ssDNA cleavage was next monitored over time in the presence of RNA targets containing no 3' PFS, a noncomplementary PFS, or a complementary (antitag) PFS. An RNA target containing noncomplementary 3' and 5' flanks (the same RNA target used in the previous experiment, Figure 7B) triggered rapid cleavage of the DNA; all DNA was cleaved within 60 seconds. However, RNA targets lacking a PFS (either with or without a 5' flank) triggered slower DNA cleavage, with less than half the DNA being cleaved after 90 seconds. Almost no DNA cleavage was observed in the presence of an RNA target containing an antitag PFS (Figure 7C). These results indicate that DNA cleavage by the *Tma*Cmr complex is partially activated by a complementary RNA protospacer, but a PFS that lacks antitag sequence is required for full activation.

The bound RNA target temporally regulates the DNase activity of Csm/Cmr complexes; cleavage and subsequent dissociation of the RNA target deactivates the DNase[22,36,37]. To determine if RNA cleavage altered the rate of DNA cleavage in our experiments, we monitored RNA cleavage by performing the same reaction with 5' end-labeled the RNA targets (rather than ssDNA). Under these conditions, where the RNA target is in ~8-fold excess of the *Tma*Cmr complex, we observed that less than 5% of each RNA target is cleaved after 60 seconds (Figure 8A-C). We also compared the rates of DNA

cleavage by wild-type *Tma*Cmr complex and by a *Tma*Cmr complex formed with the Cmr4 D26A mutant, which can bind to but cannot cleave RNA[36], and found no significant differences (Figure 8 D-F). Thus, we concluded that the difference in the rate of DNA cleavage in the presence of the different RNA targets was not due to a difference in the rate of RNA cleavage. Finally, to determine if all RNA targets can bind to the complex, Brian Learn measured the affinity of *Tma*Cmr for each target using electrophoretic mobility shift assays (EMSAs). He concluded that all targets are bound with similar affinities (apparent K_{DS} ~0.3 nM) (Figure 9). Therefore, RNA binding does not explain the observed differences in DNA cleavage. Together these data indicate that the PFS directly regulates the DNase activity of TmaCmr, likely via conformational changes in the Cas10 subunit as observed in Type III-A systems[33,53,57].

Positions -1 to -3 of the RNA target are important for regulating DNA cleavage

Data collected *in vivo* in *Pfu* have suggested a role for the specific sequence of positions -1 to -3 of the RNA target in activating Type III-B immunity[37]. To first determine which positions of the PFS regulate the DNase activity of *Tma*Cmr, we measured the extent of DNA cleavage in the presence of RNA targets containing varying complementarity to the crRNA tag. Beginning with a target containing an antitag sequence, which triggered almost no DNA cleavage (Figure 10A, Target 1), we removed complementarity one base at a time, making cumulative changes from position -8 to -1 (Figure 10A, Targets 2-8). With these targets, we observed no increase in the extent of DNA cleavage until only positions -1 and -2 contained antitag sequence (that is, positions -3 to -8 were noncomplementary), whereupon DNA cleavage was triggered to nearly the same extent as a target containing a noncomplementary PFS (Figure 10A, Targets 7 and 9). Moreover, a target with

noncomplementary sequence positions -1 to -3 triggered DNA cleavage to the same extent as the target with a fully noncomplementary PFS (Figure 10A, Target 10). These observations suggest that, like targets of *Pfu*Cmr[37], the three nucleotides of the PFS adjacent to the protospacer (positions -1 to -3 of the target) are important for regulating the DNase activity of the *Tma*Cmr complex.

The study performed in *Pfu*Cmr indicated that at least half of all possible sequences in positions -1 to -3 activated DNA cleavage[37]. To determine which sequences allow DNase activity in *Tma*Cmr, we generated a series of RNA targets containing all possible sequences in positions -1 to -3. In each of these 64 targets, the sequence of positions -4 to -8 is noncomplementary to the crRNA tag. The extent of DNA cleavage permitted by each target was measured as before. Nearly every triplet sequence allowed DNA cleavage (Figure 10B). The only sequence in positions -1 to -3 that failed to trigger DNA cleavage was the antitag sequence (GUU). DNA cleavage was reduced in the closely related AUU sequence (Figure 10B).

A guanine at position -1 of the RNA target helps to prevent DNA cleavage

In the previous experiment, a GUU sequence in positions -1 to -3 of the RNA target failed to activate DNA cleavage while UUU, CUU and AUU sequences at the same positions all activated DNA cleavage (Figure 10B), indicating the importance of position -1. However, structural data of Csm complexes shows that positions -1 of the crRNA and PFS are displaced by a loop of Csm4 (homolog of Cmr3) and therefore cannot base pair[53,57]. Together, these observations suggest that a guanine at position -1, and not base pairing, helps to prevent activation of DNA cleavage. To confirm this, we generated crRNA variants with each nucleotide at position -1 (Wild-type C-1, variants C-1U, C-1A, and C-1G). We then

compared the amount of DNA cleaved by *Tma*Cmr containing these crRNAs when activated by the above-mentioned RNA targets (those with GUU, AUU, CUU, and UUU in positions -1 to -3), thereby testing every combination of nucleotides in position -1. Regardless of crRNA sequence, and therefore base pair potential, a guanine at position -1 of the RNA target activated little DNA cleavage (Figure 11A). Conversely, robust DNA cleavage was triggered when position -1 of the target contained a cytosine, uracil, or to a lesser extent adenine (Figure 11A). All crRNA variants were functional as *Tma*Cmr bound to each was capable of robust DNA cleavage in the presence of the RNA target with uracil at position -1 (Figure 11A). These results indicate that in *Tma*Cmr, position -1 of the RNA target cannot base pair with the crRNA tag and that a guanine (the nucleotide present in repeats) at this position helps to prevent DNA cleavage.

We noted that a GUU sequence in positions -1 to -3 inhibits DNA cleavage, even when the rest of the PFS is noncomplementary to the crRNA tag (Figure 10). However, if either of the uracils in position -2 or -3 of this target is substituted for any other base, DNA cleavage is activated (Figure 10B). We wondered if substitution of the guanine in an antitag PFS would also permit DNA cleavage. We generated a series of target RNAs, beginning with a target with an antitag PFS and removing antitag sequence in a stepwise and cumulative fashion from position -1 to position -8. Upon testing these targets ability to activate DNA cleavage of *Tma*Cmr, we found that substituting a cytosine at position -1 of an otherwise complementary PFS was not enough to restore DNA cleavage, but did increase the extent of cleavage slightly (Figure 11B, Target 2). However, further removal of antitag sequence fully restored DNase activity (Figure 11B, Targets 3 to 8). We therefore conclude that a guanine at position -1 is not sufficient or necessary to prevent DNA cleavage.

Base pairing at positions -2 to -5 of the crRNA prevents DNase activation

After establishing a role for guanine in position -1 in preventing DNase activation, we wanted to investigate the role of the other positions in the 3' flank. The data presented above already indicates that positions -2 and -3 are important. Structural studies in Type III-A systems have shown that positions -2 to -5 of the crRNA tag are exposed and form base pairs with antitag sequence, but positions -1 and -6 to -8 are buried and thus cannot form base pairs[53,57]. *In vivo* data also indicate the importance of base pairing at positions -2 to -5 for regulating DNA cleavage by Csm[30,38]. To test if positions -2 to -5 are important for regulation of the DNase activity of *Tma*Cmr, we measured the extent of DNA cleavage triggered by RNA targets with varying degrees of complementarity in these positions. We found that an RNA target with a PFS complementary to only the positions buried in the Type III-A structures (-1 and -6 to -8) permitted the same extent of DNA cleavage as a target with no anti-tag sequence (Figure 12A, Target 2). RNA targets with antitag sequence at positions -2 to -5 and either guanine or cytosine at position -1 inhibited DNA cleavage (Figure 12A, Targets 3 and 4). Consistent with our earlier observations (Figure 11B), the target with a guanine at position -1 inhibited DNA cleavage to a greater extent than the target with a cytosine at the same position. We conclude that, as in Type III-A systems[30,38,53,57], positions -2 to -5 of the RNA target regulate the DNase activity of *Tma*Cmr.

To understand how much antitag sequence is needed between positions -2 and -5 to inhibit DNA cleavage, we designed RNA targets that only contain anti-tag sequence at three of these four positions. We tested these targets for their ability to activate DNA cleavage by *Tma*Cmr and found that each triggered robust DNA cleavage (Figure 12A, Targets 5-8) similar to that of an RNA target containing no antitag sequence (Figure 12A, Target 9). Thus,

if position -1 is a cytosine, all four bases in this region must have antitag sequence in order to inhibit DNA cleavage.

Given that a guanine in position -1 helps to prevent activation of DNA cleavage, we also generated a series of RNA targets containing a guanine at position -1 and one or two bases of antitag sequence in positions -2 to -5. Analysis of DNA cleavage in the presence of these targets revealed that RNAs with antitag sequence in positions -4 and -5 permitted more DNase activity than those with antitag sequence in positions -2 and -3 (Figure 12B, Targets 2-11). Thus, positions -2 and -3 are more important for blocking the DNase activity of *Tma*Cmr than positions -4 and -5.

These data implicate positions -2 to -5 of the RNA target in regulating the DNase activity of *Tma*Cmr. However, they do not address whether the identities of the nucleotides in this region are important or if these nucleotides base pair with the crRNA tag. To explore these two possibilities, we generated four variant crRNAs, each with a single nucleotide substitution in one position between -2 and -5. We then monitored DNA cleavage by the *Tma*Cmr complex containing wild-type and variant crRNAs in the presence of RNA targets that harbor compensatory substitutions in the same position, resulting in complementarity in all four positions. Robust DNA cleavage is triggered in the presence of all of these targets with wild-type crRNA (Figure 12C). In agreement with the base-pairing model, DNA cleavage was inhibited when base pairing was possible (Figure 12C). All crRNA variants were functional as each *Tma*Cmr:crRNA complex robustly cleaved DNA in the presence of an RNA target with a noncomplementary PFS (Figure 12D).

Discussion

The ability to accurately identify target sequences as foreign-derived is essential for avoiding host genome degradation in CRISPR-Cas systems. Types I, II, and V utilize specific PAM sequences to recognize foreign targets. Due to this specificity, the host genome is well protected but a point mutation in the PAM sequence of a viral target permits phage escape. Thus the foreign target identification of these systems sacrifices targeting robustness in order to prioritize host genome protection. In order to combat this issue, some systems (thus far only described in Type I) are able to acquire additional spacers from phages that have mutated their target or PAM sequence through a mechanism called primed acquisition[17,19]. By contrast, it seems that the broad targeting specificity displayed by Type III systems leads them to prioritize viral destruction over genome protection, which ensures protection from phage but carries a fitness cost[30].

The Type III Csm/Cmr complex targets RNA rather than DNA and as such does not require a PAM sequence. The PFS on the RNA target dictates whether the transcript is of host or foreign origin. Data from Type III-A *in vivo* studies demonstrate that base pairing between an antitag PFS and the crRNA tag indicates a host transcript[38], and structural studies show that this pairing deactivates the Cas10 subunit[53,57]. Alternatively, data from *in vivo* studies of a Type III-B system suggests that the sequence in the first 3 positions of the RNA target PFS indicates the origin of the transcript[37]. While it was previously unclear if these mechanisms are specific to the subtypes in which they've been described, we present evidence of a Type III-B system utilizing both mechanisms of transcript identification.

A previous study from our lab showed that DNase activation in *Tma*Cmr was achieved with an RNA target lacking a PFS[36], while studies from other systems have

indicated that a nonself PFS is required for activation[59,60]. We demonstrate that while a target lacking a PFS partially activates DNA cleavage by *Tma*Cmr, full activation requires a noncomplementary PFS (Figure 7C). With a target lacking a PFS, the rate of DNA cleavage is slower than RNA cleavage[36]; however with a nonself PFS, complete DNA cleavage occurs before RNA degradation (Figure 7C, Figure 8A, D). As RNA targets in the cell would contain PFSs, this target and the DNase activity it licenses are more biologically relevant. A more rapid rate of DNA degradation is consistent with the model that RNA cleavage and dissociation serves as a temporal control of DNase activity.

We demonstrate the importance of positions -1 to -3 of the target for self identification (Figure 10). Like in *Pfu*Cmr[37], antitag sequence in these first three nucleotides of the PFS is sufficient to deactivate the DNA cleavage of *Tma*Cmr. However, unlike *Pfu*Cmr[37], other sequences in these positions do not deactivate DNA cleavage (Figure 10B). We also showed that the nucleotide in the first position of the antitag PFS (a guanosine at position -1) is specifically recognized (Figure 11A). This is in agreement with structural models that indicate inhibition of base pairing in this position of the crRNA tag and RNA target antitag[33,53,57]. Furthermore, we present evidence that a guanosine is not required for DNA cleavage inhibition; if positions -2 to -5 contain antitag sequence, DNase activity is prohibited (Figure 11B, 12A). Like in *Sep*Csm, these positions are recognized by base pairing (Figure 12C).

These data demonstrate that *Tma*Cmr utilizes both of the mechanisms that have been proposed for Type III target identification. This represents consistency between the III-A and III-B subtypes, rather than the previously proposed difference. Originally, Type III-A systems were thought to target DNA[38,58], while Type III-B systems were proposed to target

RNA[54–56]; however, later work established a unified model by demonstrating that both Csm and Cmr degrade DNA and RNA[22,36,37]. Perhaps the data presented in this study support another unified model for Type III systems; it is conceivable that effector complexes from all Type III subtypes are able to use both base pairing and protein recognition for target identification. The experimental setups of the two studies that proposed the conflicting mechanisms[37,38] may have biased the results and therefore allowed the researchers to overlook this possibility. Alternatively, perhaps target identification falls on a spectrum. That is, some Type III effectors may recognize antitag targets through base pairing alone (*e.g. Sep*Csm[38]) while others may recognize these based on sequence specificity (*Pfu*Cmr[37]), and yet others utilize both mechanisms (*Tma*Cmr) to different extents. Further work in more species and subtypes is required to understand the prevalence of these two mechanisms.

Figures

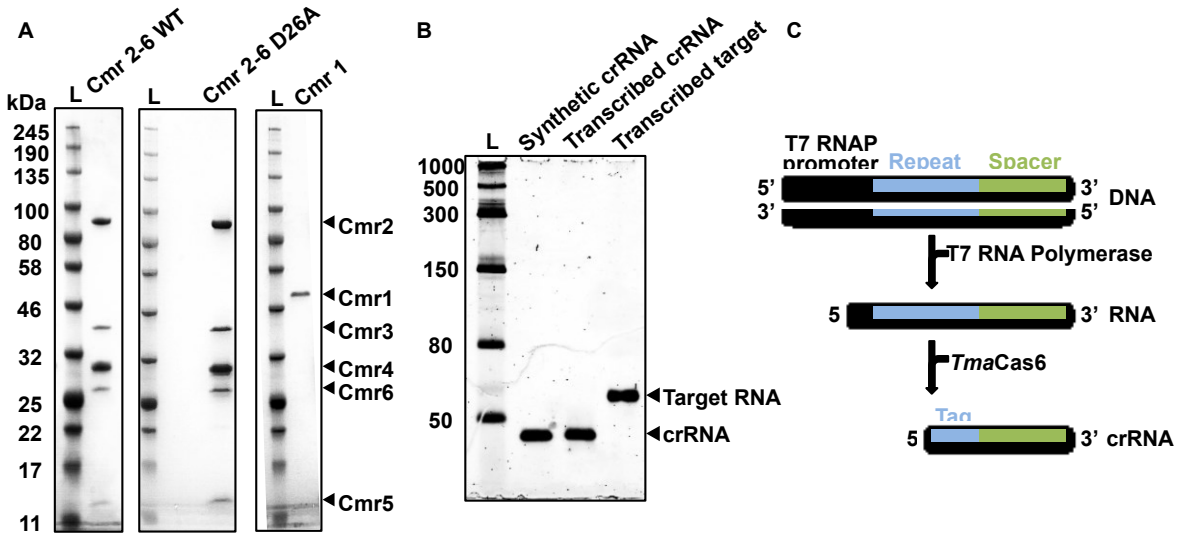

Figure 6. *Tma*Cmr protein and RNA generation. (A) SDS-PAGE gels of recombinantly purified *Tma*Cmr2-6 WT (left), *Tma*Cmr2-6 D26A (middle), and *Tma*Cmr1 (right). L indicates ladder (NEB P7712). (B) Urea-PAGE gel of crRNA and target RNA. L indicates ladder (NEB N0364). Synthetic crRNA was included as a control to confirm appropriate size after crRNA transcription and processing by *Tma*Cas6. (C) Schematic of crRNA production. A double-stranded DNA template was first transcribed by T7 RNA Polymerase and the resulting RNA was processed by *Tma*Cas6 to produce a mature crRNA consisting of an 8 nt tag and 37 nt spacer.

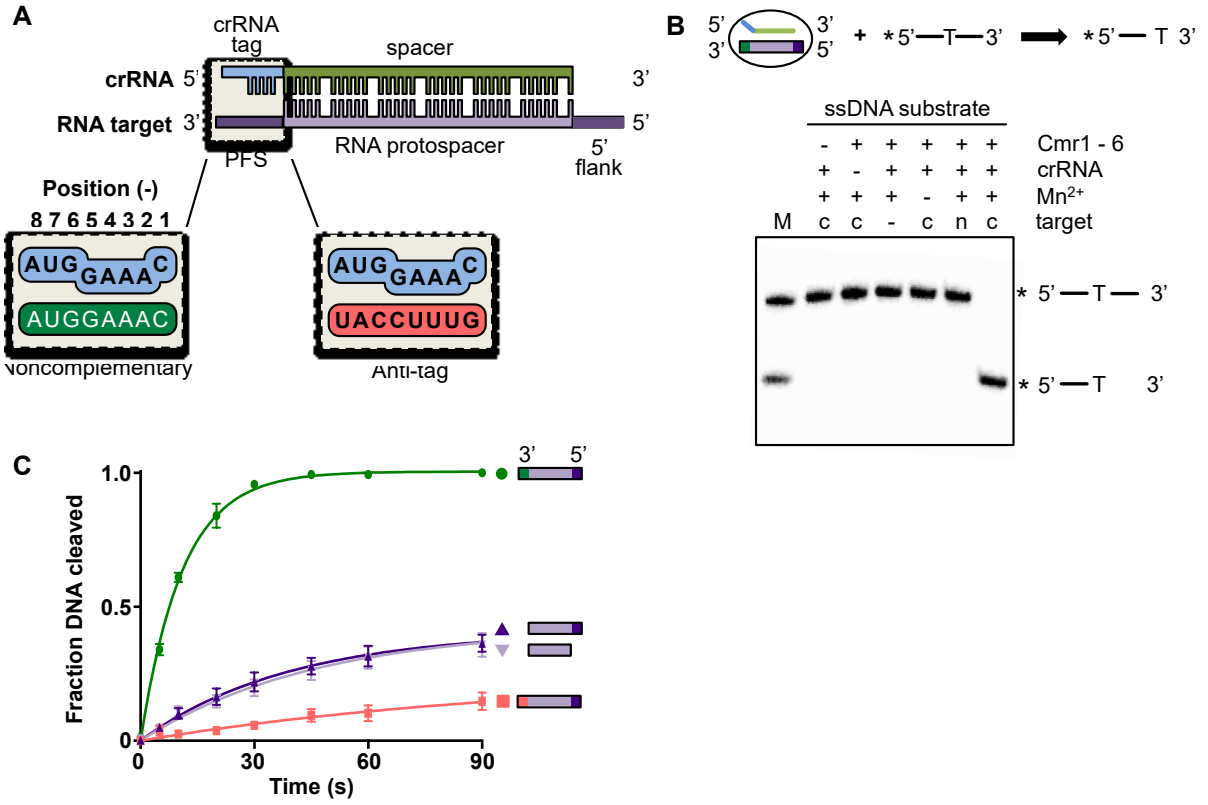

Figure 7. DNA cleavage by *Tma*Cmr is transcript-dependent. (A) DNA cleavage in Type III CRISPR-Cas systems is activated by a transcript with a 3' protospacer flanking sequence (PFS) that lacks complementarity to the crRNA (shown in green). RNA targets containing an antitag PFS (shown in pink) fail to activate DNA cleavage. Structural studies indicate that positions -2 to -5 of the crRNA are available for base pairing. (B) Urea-PAGE analysis of DNA cleavage by *Tma*Cmr. First lane (M) contains markers of substrate and expected product. 5' radiolabeled DNA was incubated with (+) or without (-) *Tma*Cmr 1-6, crRNA, Mn^{2+}, and target RNA for 60 seconds at 80° C. Target "c" and "n" indicate RNA protospacers complementary or noncomplementary to the crRNA spacer region, respectively. (C) Quantification of DNA cleavage by *Tma*Cmr complex over time upon activation with RNA targets containing various 3' and 5' flanks. RNA targets represented by green circles have a noncomplementary PFS, while purple triangles represent targets without a PFS (dark purple includes a 5' flank, light purple lacks a 5' flank), and pink squares represent targets with an antitag PFS.

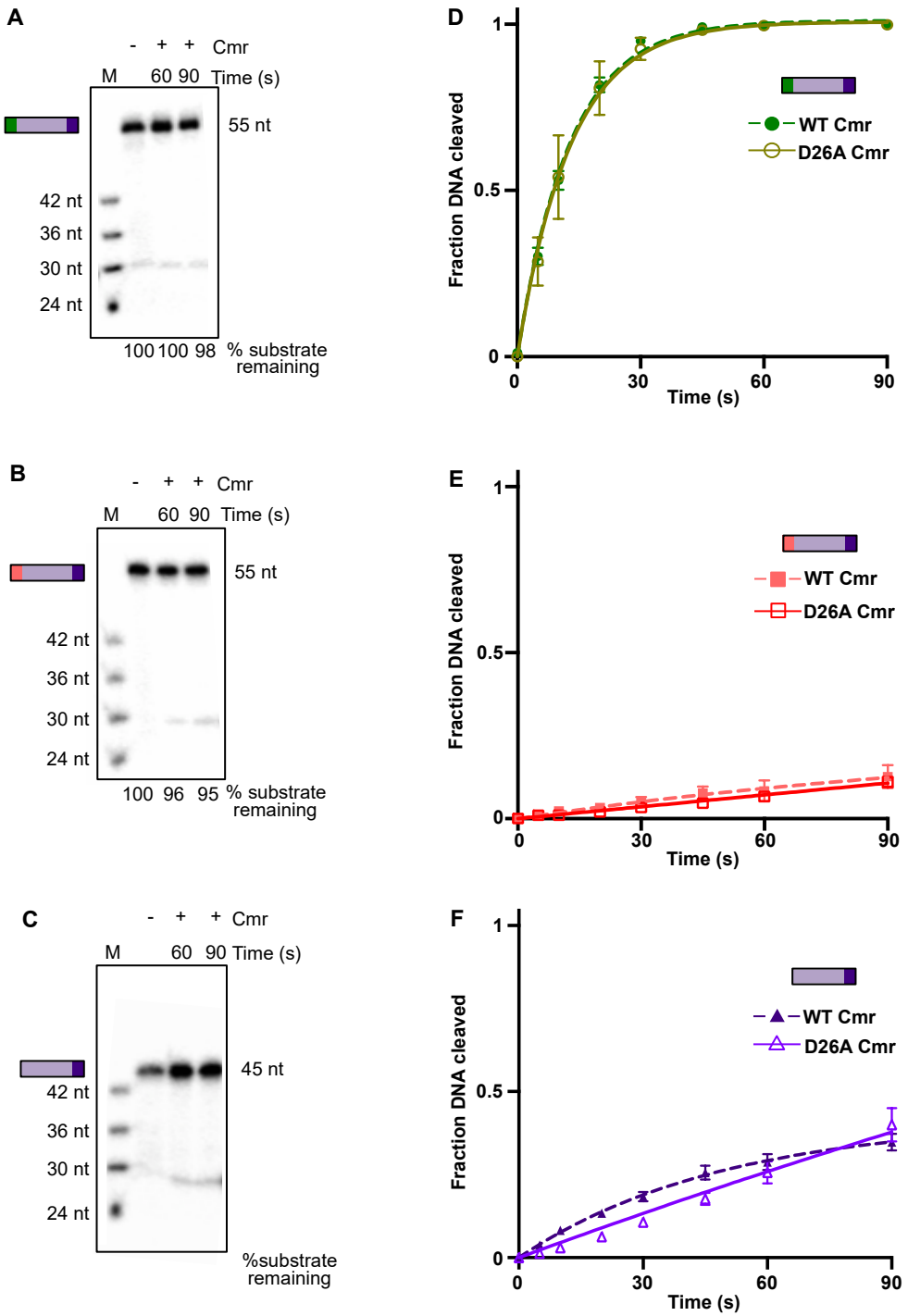

Figure 8. Verification of target RNA persistence. Green (panels A, D) indicates a noncomplementary PFS, pink (panels B, F) indicates an antitag PFS, and purple (panels C, E) indicates no PFS. (A, B, C) Representative Urea-PAGE showing persistence of RNA target after incubation with *Tma*Cmr for 60 and 90 seconds under DNA cleavage conditions. M indicates marker generated by 5' radiolabeling expected cleavage products (synthetic) of a 55 nucleotide substrate. (D, E,F) Quantified time courses of DNA cleavage with wild-type (dashed lines and filled shapes) and RNA cleavage mutant D26A (solid lines, open shapes) *Tma*Cmr complexes upon activation with targets with various PFSs.

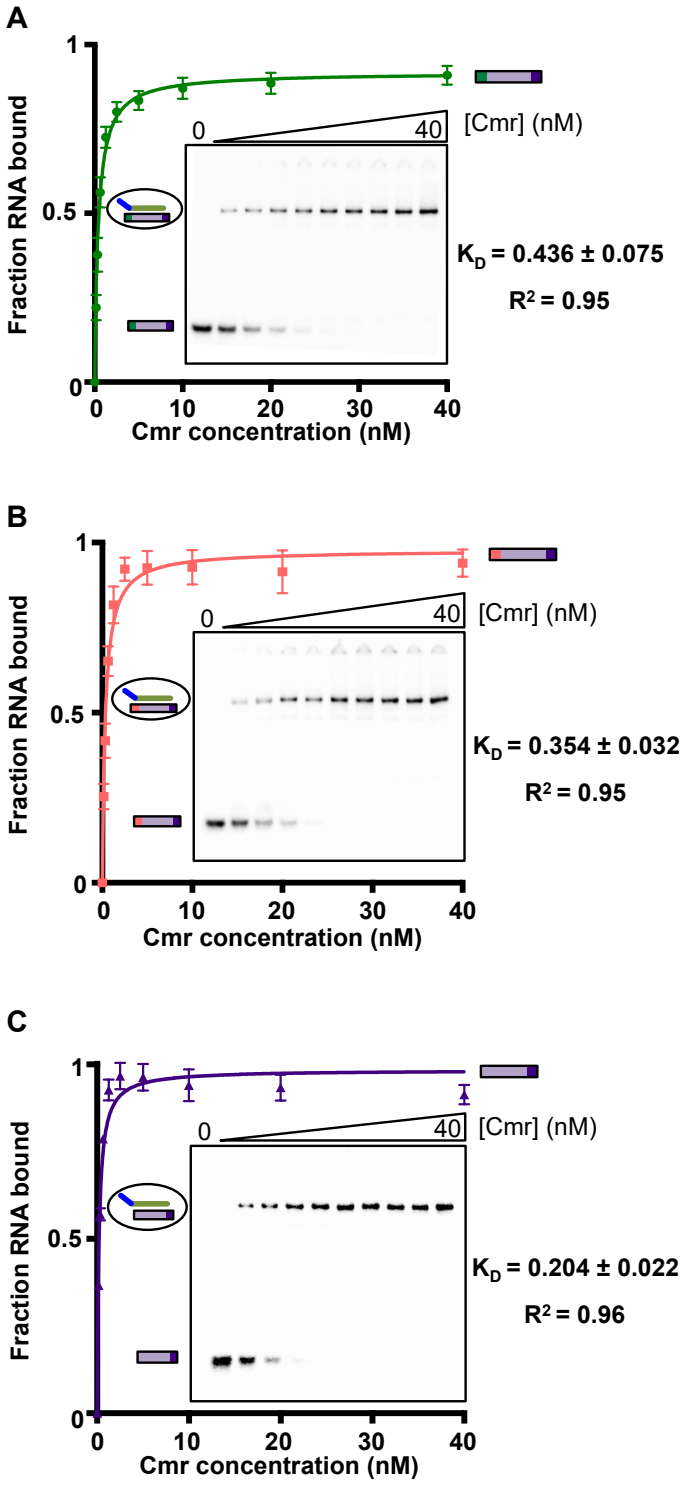

Figure 9*. Verification of RNA target binding by *Tma*Cmr. Binding curves with representative EMSA and binding constants of *Tma*Cmr with RNA targets with the following PFS. (A) noncomplementary, (B) antitag, (C) no PFS. *All Figure 9 experiments and quantification were performed by Brian Learn.

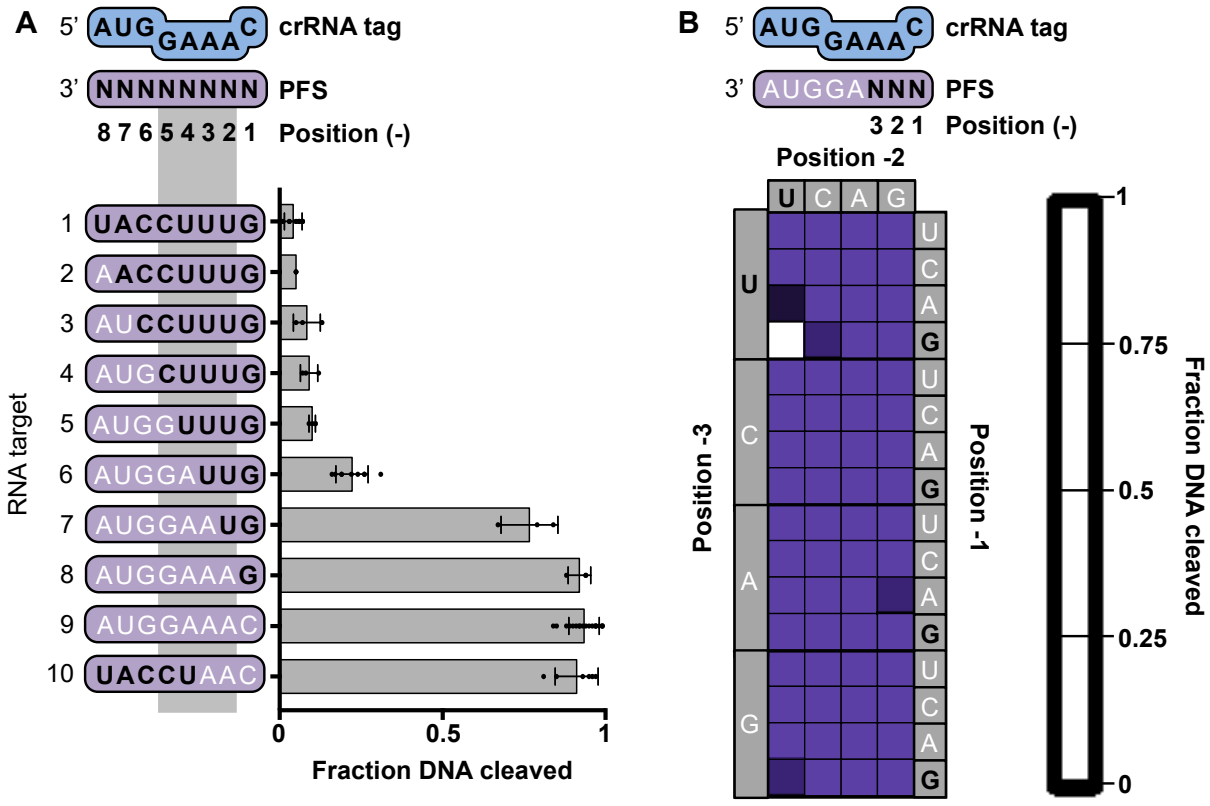

Figure 10. Importance of RNA target positions -1 to -3 in self target identification. (A) Quantification of DNA cleavage by *Tma*Cmr complex when activated by RNA targets with various positions of antitag sequence in the PFS. Black letters indicate positions of antitag sequence. (B) Heat map of *Tma*Cmr DNA cleavage upon activation with RNA targets with all sequences in positions -1 to -3. All PFS contain noncomplementary sequence in positions -4 to -8. Dark purple indicates a high level of activation, while white indicates low activation.

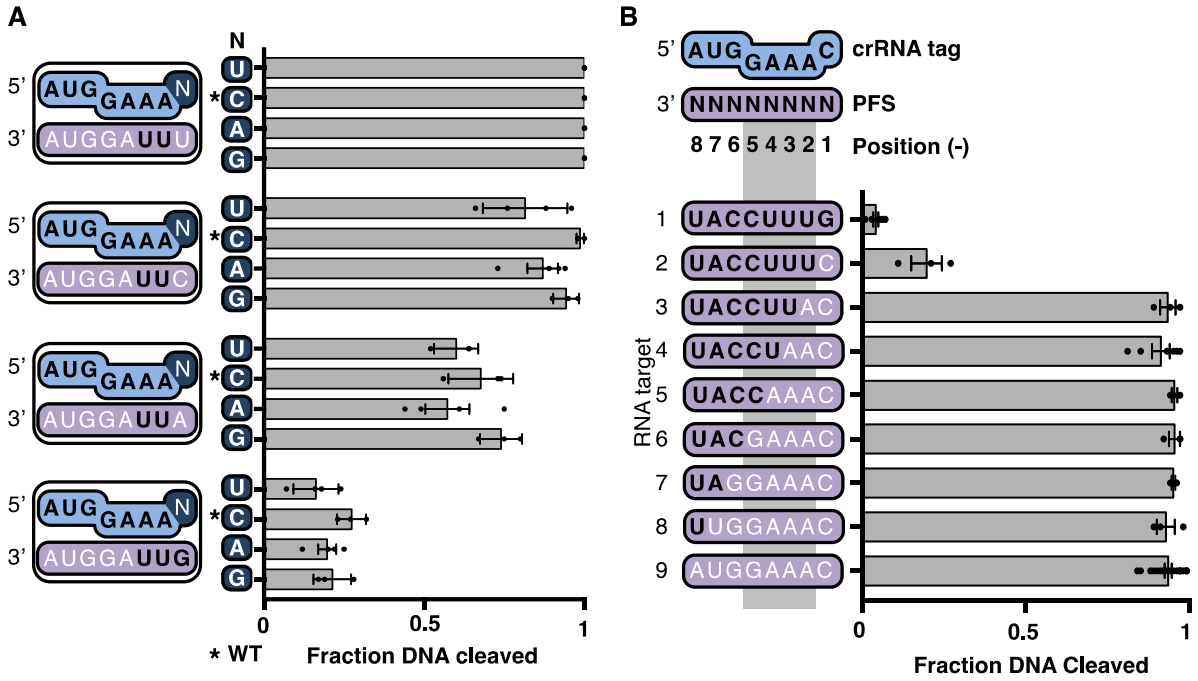

Figure 11. Role of RNA target position -1 in target identification. (A) Quantification of DNA cleavage by *Tma*Cmr complex. Position -1 WT or mutant crRNAs were paired with RNA targets containing U, C, A, and G at position -1 and DNA cleavage was quantified. All targets contain the same sequence of 5'-UUAGGUA-3' in positions -2 to -8. Black letters indicate positions of antitag sequence. (B) Quantification of DNA cleavage by *Tma*Cmr complex when activated by transcripts containing PFSs with various positions of antitag sequence. Black letters indicate positions of antitag sequence.

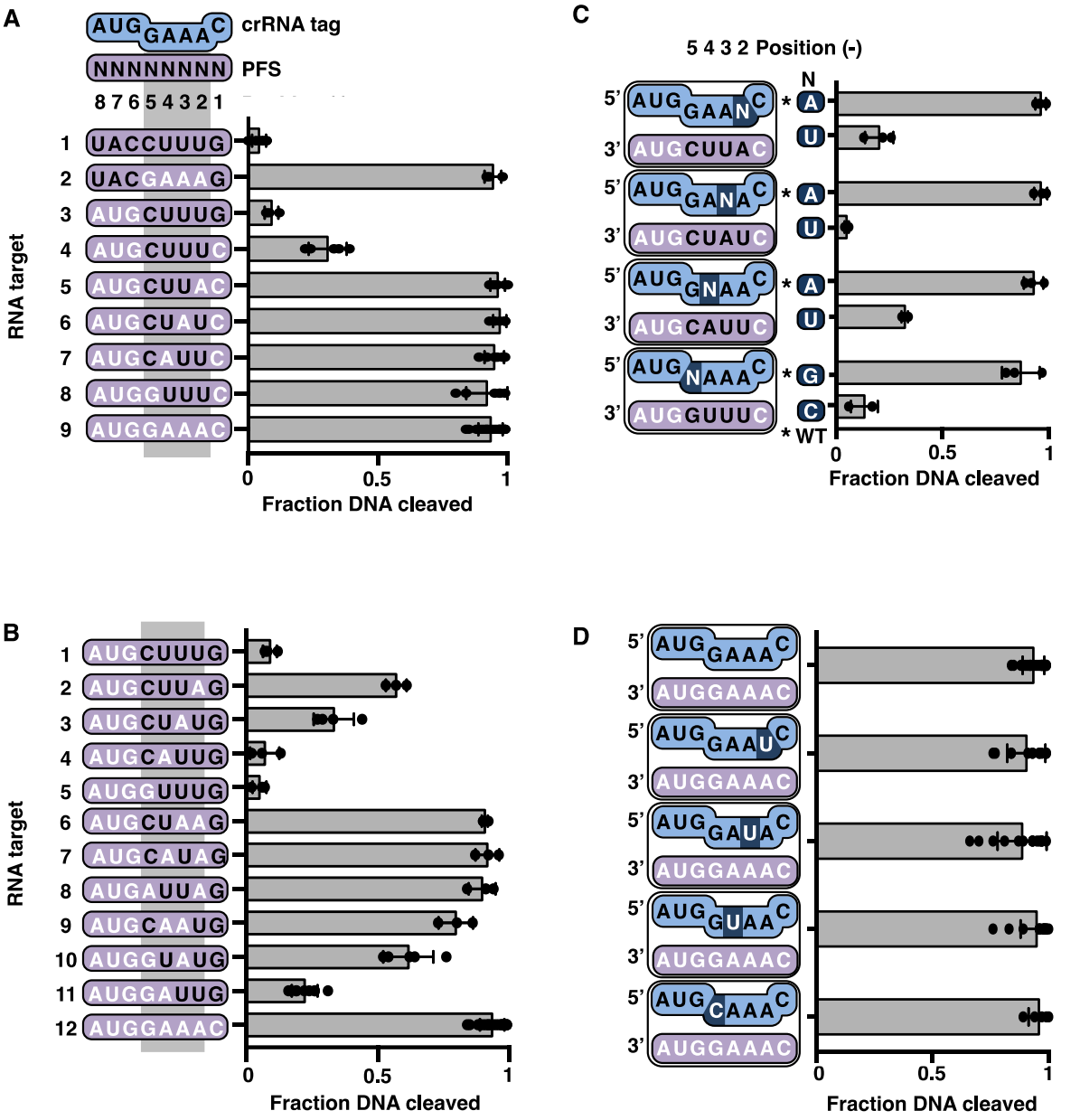

Figure 12. Role of target positions -2 to -5 in self target identification. (A,B) Quantification of DNA cleavage by Cmr complex when activated by RNA targets with various positions of antitag sequence. Black letters indicate antitag sequence. (*C*) Quanitification of DNA cleavage by *Tma*Cmr complex. Target RNAs with antitag sequence in three of four positions from -2 to -5 were paired with WT crRNA (*) or a crRNA mutated such that the target gained full complementarity in this region. (D) Quantification of DNA cleavage by *Tma*Cmr in complex with WT and mutant crRNAs.

Materials and Methods

Expression and purification of recombinant proteins

TmaCmr2-6 complex was made by expressing each subunit individually and pooling all cells for purification. *Tma*Cmr3 (codon-optimized), *Tma*Cmr4 (wild-type or D26A), and *Tma*Cmr5 were previously individually cloned into pRSFDuet-1 and *Tma*Cmr6 was previously cloned into pHAT2[36]. *Tma*Cmr2 was codon-optimized and cloned into pRSFDuet-1 (Novagen) as well. These vectors were transformed individually into T7Express cells (NEB). 1 L of each (2L *Tma*Cmr6) was grown with appropriate antibiotics at 37 °C to OD600 = 0.4 in Luria-Bertani (LB) media followed by induction with 200 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) and overnight growth at 20 °C. Cells were pelleted and resuspended together in lysis buffer [1M KCl, 20 mM tris pH 8.0, 10 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine (TCEP)] with 1 μM E-64, 0.1 μM phenylmethylsulfonyl fluoride (PMSF), 1.7 μM bestatin, 2.5 μM pepstatin A, and 1 mM each ATP and MgCl2. The cells were then lysed using a microfluidizer. The lysate was heat treated in an 80 °C water bath for ten minutes, then clarified by centrifugation. This sample was passed over a 5 mL immobilized metal affinity chromatography (IMAC) column (Bio-Rad) charged with nickel sulfate and equilibrated with lysis buffer. After washing with 20 column volumes (CV) lysis buffer, the sample was eluted [500 mM KCl, 20 mM tris pH 8.0, 250 mM imidazole, 1 mM TCEP] and further purified by size exclusion (HiLoad 26/60 S200 column, GE Healthcare) in gel filtration buffer [350 mM KCl, 20 mM tris pH 8.0, 1 mM TCEP] (Figure 6A).

Cmr1 (Figure 6A) and Cas6 were cloned into pHAT2 and purified separately as described above without heat treatment.

His_6-P266L T7 RNA polymerase[62] was expressed in T7Express cells. 1L was grown to OD600 = 0.3 at 37 °C, induced with 200 μM IPTG, and grown overnight at 20 °C. The cells were pelleted and resuspended in T7RNAP lysis buffer [250 mM NaCl, 10 mM imidazole, 20 mM HEPES pH 7.0, 10 mM dithiothreitol (DTT), and 5% glycerol] with 1 μM E-64, 0.1 μM PMSF, 1.7 μM bestatin, 2.5 μM pepstatin A. Cells were lysed using a microfluidizer and clarified by centrifugation. The sample was added to 3 mL of nickel-charged Profinity™ IMAC resin (Bio-Rad) in T7RNAP lysis buffer and gently rocked at 4 °C for 30 minutes. The supernatant was removed and the resin was washed with 25 CV lysis buffer. The protein was then eluted [250 mM NaCl, 250 mM imidazole, 20 mM HEPES pH 7.0, 10 mM DTT, and 5% glycerol].

RNA generation

All RNAs under 45 nucleotides were purchased (Sigma), while longer RNAs were synthesized by *in vitro* transcription. 5μM complementary DNA templates (Sigma) were annealed [50 mM NaCl, 10 mM tris pH 8.0, and 1 mM ethylenediaminetetraacetic acid (EDTA)] by slow cooling from 95 °C to room temperature. 100 nM dsDNA was then incubated with 5 mM each rNTP, 15 mM $MgCl_2$, 0.1 mg/mL P266L T7 RNA polymerase, and transcription mix [25 mM tris pH 8.0, 2 mM spermidine, 40 mM DTT] for 2 – 3 hours at 37 °C.

crRNAs were transcribed as described above. Following transcription, RNAs were treated with Cas6 to trim the 5' repeat sequence: the transcription reaction was incubated

with 200 nM Cas6 for 20 minutes at 80 °C in Cas6 cleavage buffer [65 mM KCl, 20 mM HEPES pH 7.0, 20 mM EDTA, and 10% glycerol] (Figure 6B-C).

Transcripts (including processed crRNAs) were then gel extracted, ethanol precipitated, and resuspended in RNA storage solution (Thermo Fisher) with 1 unit Ribolock RNase inhibitor (Figure 6B).

Radiolabeling of oligonucleotides

DNA substrate (Sigma) and synthetic RNAs were gel purified and radiolabeled with 1-2 pmol [γ32P]ATP (Perkin-Elmer) by incubating 200 nM DNA in 1x T4 polynucleotide kinase (PNK) buffer with 10 units T4 PNK (NEB) at 37 °C for 30 minutes, followed by heat inactivation at 65 °C for 20 minutes.

Prior to radiolabeling, 5' ends of transcripts were dephosphorylated by incubating 200 nM RNA in 1x CutSmart buffer with 1 unit shrimp alkaline phosphatase (rSAP) (NEB) for 30 minutes at 37 °C, followed by a 20 minute heat inactivation at 75 °C. 10 units T4 PNK, 1x PNK buffer and 2 pmol [γ32P]ATP were added and the reaction incubated at 37 °C for 30 minutes, followed by heat inactivation for 20 minutes at 65 °C. The transcripts were then gel extracted, ethanol precipitated, and resuspended in RNA storage solution.

DNA cleavage assays

*Tma*Cmr complex was first formed by incubating crRNA, *Tma*Cmr1, and *Tma*Cmr2-6 at 80 °C for 20 minutes in reaction buffer [100 mM KCl, 50 mM HEPES pH 7.0, 1 mM TCEP, and 1 mM $MnCl_2$]. 25 nM *Tma*Cmr complex was used in cleavage reactions. Following complex formation, 95 nM unlabeled DNA substrate and 5 nM 5' radiolabeled DNA were added. Reactions were initiated upon addition of 200 nM RNA (or RNA storage

solution in no transcript controls) and quenched [90% formamide, 2.5% glycerol, 0.01% SDS, 0.01% bromophenol blue, 0.01% xylene cyanol, and 1 mM EDTA] after 1 minute unless otherwise indicated. These samples were then run on 15% polyacrylamide urea gels and visualized by phosphorimaging (FujiFilm FLA-7000). Images were quantified using Image Gauge (FujiFilm) and data analysis was done in Prism (GraphPad Software) and Excel (Microsoft). All data points are the average of at least three replicates and error bars represent standard error of the mean.

RNA cleavage assays

RNA cleavage assays were performed with the same protocol as described above for DNA cleavage but with 5 nM 5' radiolabeled target RNA, 195 nM unlabeled target RNA, and 100 nM unlabeled DNA. Synthetic RNAs of expected products were radiolabeled and used as markers to confirm appropriate product size.

CHAPTER 3: TARGET MISMATCH TOLERANCE IN A TYPE III-B CRISPR-CAS SYSTEM

Type III CRISPR-Cas systems offer robust protection from phage by utilizing a CRISPR-RNA (crRNA)-guided effector complex to degrade foreign RNA and DNA. These systems are more robust than other CRISPR-Cas types in part because they are activated by more target PFS sequences, as described in Chapter 2. Additionally, they seem to tolerate more mismatches. For example, unlike other types of CRISPR-Cas interference, duplex formation between the crRNA and the target RNA does not require a seed sequence in Type III systems; mismatches are tolerated across the entire spacer:protospacer duplex. While other types typically tolerate fewer than four mismatches, a study of *Sep*Csm demonstrated that many mismatches do not affect Type III-A immunity and only with extensive mismatching (five or more nucleotides) in the PFS-adjacent segment of the target is immunity weakened. It is unknown if this level of mismatch tolerance is universal among all Type III systems, or if it is subtype- or species-specific.

Here, we present support for high mismatch tolerance among Type III systems. Utilizing recombinantly purified Cmr complex and RNA targets generated through *in vitro* transcription, we studied the requirement for base pairing between segments of the crRNA spacer and the target protospacer. We find that, like the *Sep*Csm complex, *Tma*Cmr activation is significantly reduced by a target that contains five mismatched nucleotides at the PFS-adjacent segment (segment 1). We also provide evidence that a target activates DNA cleavage even with significant regions of mismatching outside of segment 1 as long as it is

still able to bind to the *Tma*Cmr complex. These data support findings in Type III-A systems and extend the defining characteristic of mismatch tolerance to Type III-B.

Introduction

CRISPR-Cas interference complexes are guided by their crRNAs to complementary nucleic acid targets. Activation of the effector relies upon complementarity between the target sequence and the crRNA as well as identification of the target as foreign and not host-derived. CRISPR-Cas Types I, II, and V rely upon the use of a PAM sequence for foreign nucleic acid identification[15]; however, as discussed in Chapter 1, Type III systems rely on the lack of antitag sequence in the target PFS for self target identification and are therefore activated by more target sequences. This reduces the potential for viral escape mutants in Type III systems[30].

CRISPR-Cas types require different levels of complementarity between the target protospacer and crRNA spacer sequences. The effectors of Types I, II, V, and VI have been shown to rely upon a seed sequence for duplex formation and any mismatches in this region prevent further base pairing (and therefore complex activation)[10]. Outside of the seed region, mismatches are often tolerated to an extent, particularly in the PAM-distal region of the protospacer. Cascade (Type I) is able to target DNA containing four or fewer mismatched positions outside of the seed region[63]. Cas9 (Type II) can tolerate only one or two mismatches, perhaps due in part to fewer positions of the crRNA being available for base pairing to begin with[64]. Cas12a (Type Va) is sensitive to single mismatches in the PAM-proximal end but can tolerate double mismatches in the PAM-distal end[40]. Cas13a (Type VIa) tolerates one or two mismatches, depending on the position[40]. Further mismatching between the target and crRNA allow for viral escape in each of these cases.

Type III systems likely do not require a seed sequence; data on the *Sep*Csm (Type III-A) system indicate that mismatches are tolerated in all regions of the protospacer[21,30]. In fact, extensive accumulation of mismatches is tolerated throughout the protospacer in *Sep*Csm, and only with contiguous mismatches in the first five nucleotides of the protospacer is immunity weakened[30]. Escape mutants were undetectable (below the limit of detection, one in 10^{10}) if an essential gene was targeted by Type III-A, while one in 10^9 viral particles escaped Type II-A immunity[30]. Deletion of the targeted region seems to be the only escape from Type III-A systems, which is rarely possible as targeted phage genes are typically essential; thus Type III-A CRISPR-Cas can cause extinction of a targeted phage[30].

It is unclear if tolerance for target mismatches is universal for all subtypes of Type III systems. We were particularly curious if Type III-B systems are capable of cleaving DNA given mismatched RNA targets. To address this, we generated RNA targets containing various levels of complementarity to the crRNA and tested their ability to activate recombinantly purified *Tma*Cmr. Crystal structures of Cmr have revealed that, as in Cascade, a loop is inserted between every sixth nucleotide of the crRNA and target RNA duplex, flipping out the bases at these positions[25,53,57]. We defined each of these five-base pair regions as a segment and introduced changes in the target RNA at one segment at a time. We found that all segment mismatches were tolerated except segment 1 (the PFS-proximal region). However, contiguous mismatching in this segment is required to inactivate DNA cleavage by *Tma*Cmr, supporting the notion that Type III systems do not require a seed[30]. We also demonstrate that outside of segment 1, extensive mismatching is tolerated so long as the target maintains enough complementarity to bind to the complex.

Results

Segments of five mismatches are tolerated except in the PFS-proximal segment

Immunity in Type III systems seems to be highly tolerant of mismatches between the crRNA and RNA protospacer[30]. The structure of the crRNA:RNA target duplex in Type III complexes is arranged into segments of five base pairs separated by disrupted one base gaps[25,53,57]. We generated RNA targets (via *in vitro* transcription) in which we introduced five nucleotide blocks of mismatches at each of these segments in order to determine where mismatches might be tolerated for DNA cleavage by recombinantly purified *Tma*Cmr complex (Figure 7). We monitored DNA cleavage by *Tma*Cmr in the presence of these RNA targets and found that mismatches within the PFS-proximal segment 1, which includes the five nucleotides at the most 3' end of the protospacer (the most 5' end of the crRNA spacer) prevented DNA cleavage (Figure 13A, Target 2) However, an RNA target containing any other mismatched segment fully activated DNA cleavage (Figure 13A, Targets 3-7). Brian Learn performed EMSAs to determine that the RNA target with mismatched segment 1 bound to *Tma*Cmr with the same affinity as a fully complementary target (Figure 9) and a target with a mismatched segment 6 (Figure 13C). The RNA target itself was cleaved to the same extent (<5% after 60 seconds, Figure 13D) as a fully complementary target (Figure 8A) We conclude that base pairing in segment 1 of a bound RNA target is important for the activation of DNA cleavage.

To see how many base pairs are needed for DNase activation, we also monitored DNA cleavage in the presence of RNA targets containing single point or accumulating mismatches within segment 1, finding that as few as one to two base pairs in this segment is sufficient to activate DNA cleavage (Figure 14). These observations show that base pairing

in segment 1 is necessary to activate DNA cleavage, but every position need not be base paired for activation.

Mismatching in other segments is tolerated until target no longer binds

To determine how many mismatched segments are tolerated for DNA cleavage activation, we tested RNA targets containing an increasing number of mismatched segments, starting at segment 6 (the 5' end of the RNA protospacer). We determined that as the number of mismatched segments increases, DNA cleavage decreases. When only segments 1 and 2 remain complementary, DNA cleavage does not proceed above background (DNA cleavage in the presence of a fully mismatched RNA protospacer) (Figure 13A). Brian Learn again performed EMSAs to assess binding of *Tma*Cmr to these targets, as the complex is likely to have a lower affinity for targets that contain extensive mismatches to the crRNA. He showed that targets which failed to activate DNA cleavage were also poorly bound by *Tma*Cmr (Figure 13 C). We therefore concluded that as long as a target contains a complementary segment 1 and can bind to the *Tma*Cmr complex, it will activate DNA cleavage.

Discussion

Most CRISPR-Cas systems are highly specific for targets containing full complementarity to the crRNA spacer. Most importantly in Types I, II, V, and VI is complementarity to the seed region; a lack of base pairing between the crRNA and target within this seed region prevents effector activation in these types[10]. Outside of this seed, these types are typically only able to tolerate between one and four mismatches[40,63,64]. A lack of a seed sequence has been suggested for some Type III systems (with evidence presented from *Sep*Csm), in which extensive mismatches are tolerated in all regions of the protospacer;

in fact, this system is so tolerant of mismatches that the most probable way for a phage to escape targeting is to delete the target from its genome[30]. Evidence from *Sep*Csm suggests that some amount of base pairing in the first ten nucleotides nearest to the PFS and crRNA tag is important for immunity; yet an effect on immunity is only detected with five or more mismatches in this region[30]. However, there is insufficient evidence to understand if extensive mismatch tolerance is a universal characteristic of Type III targeting.

Here we demonstrate that the *Tma*Cmr complex also tolerates extensive mismatching. Like *Sep*Csm[30], base pairing is most important in the segment nearest the PFS but DNA cleavage is still activated with targets that contain up to four of five mismatches in this first segment (Figure 13A, Figure 14). Although a target with mismatched segment 1 does not activate DNA cleavage, this target is still bound (Figure 13B); thus this segment does not constitute a seed region in which complementarity is required for binding. We also demonstrate that mismatched nonself targets activate DNA cleavage as long as the protospacer maintains enough complementarity (including segment 1) to bind to the complex (Figure 13A, B). This supports the finding that phage mutants are unable to escape from Type III-A targeting by *Sep*Csm[30] and suggests that this may be a common feature of Type III systems.

The findings presented in Chapters 2 and 3 demonstrate that *Tma*Cmr exhibits broad targeting capabilities and that this may be a common feature of all Type III systems. Whereas Types I and II often must acquire multiple spacers for each target for effective interference, Type III systems only require one spacer[30]. The broad targeting capabilities of Type III systems ensures phage elimination but has been demonstrated to transfer a fitness cost to the host[30]. Spacer acquisition has been proposed to occur less frequently in Type III

systems than in other CRISPR-Cas systems, perhaps in part due to this fitness cost[30]. Additionally, acquisition has not yet been detected in Type III systems, and acquisition machinery is absent in many Type III-B loci[28]. Thus it is unclear how spacers are acquired, but evidence suggests that once acquired, these spacers provide Type III CRISPR-Cas with a robust memory, even upon phage mutation.

Figures

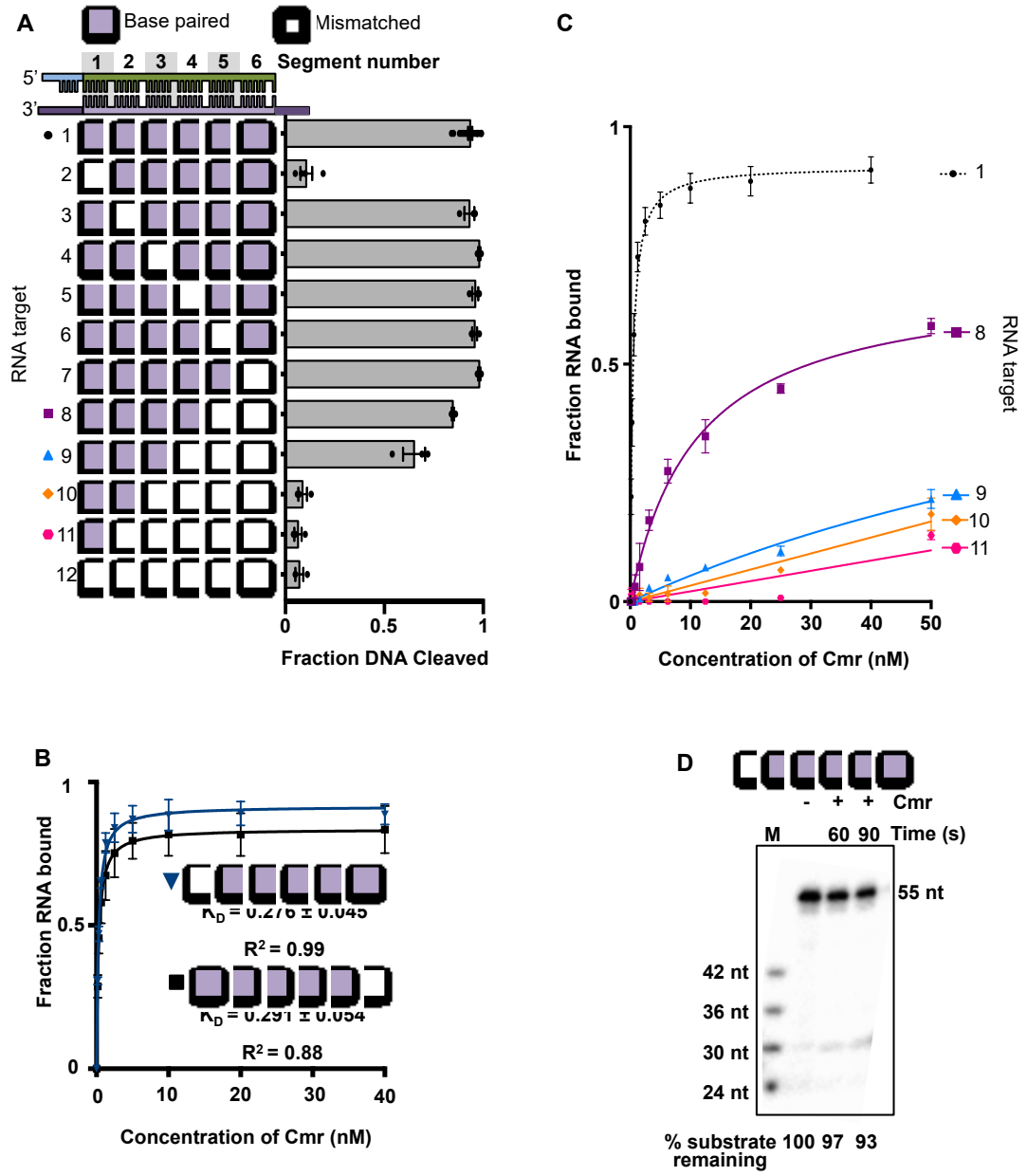

Figure 13. Tolerance of protospacer mismatches. (A) Quantification of DNA cleavage by Cmr complex activated by targets containing protospacers that are mismatched to the crRNA in one or multiple segments. Segments are defined by the five-nucleotide stretches of base pairs within the structure of the crRNA-target duplex. All targets contain a noncomplementary PFS. (B, C) Quantification of Cmr complex binding to mismatched targets (B) 2 and 7 and (C) 8, 9, 10, and 11 from panel A. These experiments (Panels B and C) and quantifications were performed by Brian Learn. (D) Representative Urea-PAGE showing persistence of RNA target after incubation with *Tma*Cmr for 60 and 90 seconds under DNA cleavage conditions. M indicates marker generated by 5' radiolabeling expected cleavage products (synthetic) of a 55 nucleotide substrate.

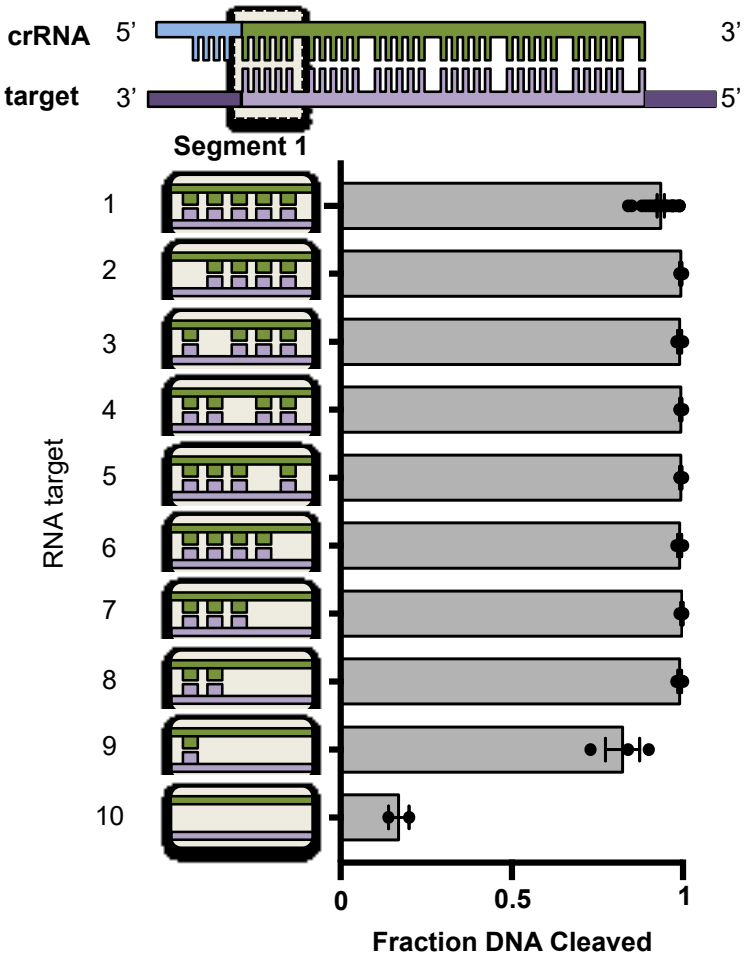

Figure 14. Tolerance of single or multiple mismatches in segment 1. Quantification of DNA cleavage by *Tma*Cmr upon activation with targets containing one or more mismatches within segment 1.

Materials and Methods

All materials and methods as described in Chapter 2.

REFERENCES

1. Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient Virus World and evolution of Cells. *Biol. Direct* **1**, (2006).
2. Comeau, A. M. *et al.* Exploring the prokaryotic virosphere. *Res. Microbiol.* **159**, 306–313 (2008).
3. Azam, A. H. & Tanji, Y. Bacteriophage-host arm race : An update on the mechanism of phage resistance in bacteria and revenge of the phage with the perspective for phage therapy. *Appl. Microbiol. Biotechnol.* (2019). doi:10.1007/s00253-019-09629-x
4. Van Valen, L. A new Evolutionary Law. *Evol. Theory* **1**, 1–30 (1973).
5. Stern, A. & Sorek, R. The phage-host arms-race: Shaping the evolution of microbes. *Bioessays* **33**, 43–51 (2011).
6. Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu. Rev. Microbiol.* **71**, 233–261 (2017).
7. Dy, R. L., Richter, C., Salmond, G. P. C. & Fineran, P. C. Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annu. Rev. Virol.* **1**, 307–331 (2014).
8. Willkomm, S., Makarova, K. S. & Grohmann, D. DNA silencing by prokaryotic Argonaute proteins adds a new layer of defense against invading nucleic acids. *FEMS Microbiol. Rev.* **42**, 376–387 (2018).
9. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science (80-.).* **315**, 1709–1713 (2007).
10. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).
11. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide Sequence of the iap Gene , Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia coli , and Identification of the Gene Product. *J. Bacteriol.* **169**, 5429–5433 (1987).
12. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
13. Mojica, F. J. M., Díez- Villaseñor, C., García-Martínez, J. & Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J. Mol. Evol.* **60**, 174–182 (2005).
14. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA , and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).

15. Koonin, E. V, Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).

16. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).

17. Mcginn, J. & Marraffini, L. A. Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat. Rev. Microbiol.* **17**, 7–12 (2019).

18. McGinn, J. & Marraffini, L. A. CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. *Mol. Cell* **64**, 616–623 (2016).

19. Gleditzsch, D. *et al.* PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol.* **00**, 15476286.2018.1504546 (2018).

20. Meeske, A. J. & Marraffini, L. A. RNA Guide Complementarity Prevents Self-Targeting in Type VI CRISPR Systems. *Mol. Cell* **71**, 791–801 (2018).

21. Pyenson, N. C. & Marraffini, L. A. Type III CRISPR-Cas systems: when DNA cleavage just isn' t enough. *Curr. Opin. Microbiol.* **37**, 150–154 (2017).

22. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č. & Siksnys, V. Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Mol. Cell* **62**, 295–306 (2016).

23. Terns, M. P. CRISPR-Based Technologies: Impact of RNA-Targeting Systems. *Mol. Cell* **72**, 404–412 (2018).

24. Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science (80-.).* **353**, (2016).

25. Mulepati, S., Héroux, A. & Bailey, S. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science (80-.).* **345**, 1479–1484 (2014).

26. Jackson, R. N., van Erp, P. B., Sternberg, S. H. & Wiedenheft, B. Conformational regulation of CRISPR-associated nucleases. *Curr. Opin. Microbiol.* **37**, 110–119 (2017).

27. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 (2008).

28. Tamulaitis, G., Venclovas, Č. & Siksnys, V. Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends Microbiol.* **25**, 49–61 (2017).

29. Zhu, Y., Klompe, S. E., Vlot, M., van der Oost, J. & Staals, R. H. J. Shooting the messenger: RNA-targetting CRISPR-Cas systems. *Biosci. Rep.* **38**, BSR20170788 (2018).

30. Pyenson, N. C., Gayvert, K., Varble, A., Elemento, O. & Marraffini, L. A. Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. *Cell Host Microbe* **22**, 343–353 (2017).

31. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science (80-.).* **357**, 605–609 (2017).

32. Niewoehner, O. *et al.* Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).

33. Wang, L. *et al.* Dynamics of Cas10 Govern Discrimination between Self and Nonself in Type III CRISPR-Cas Immunity. *Mol. Cell* **73**, 278–290 (2019).

34. Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V. & Aravind, L. CARF and WYL domains: Ligand-binding regulators of prokaryotic defense systems. *Front. Genet.* **5**, 1–9 (2014).

35. Rouillon, C., Athukoralage, J. S., Graham, S., Grüschow, S. & White, M. F. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *Elife* **7**, 1–22 (2018).

36. Estrella, M. A., Kuo, F. T. & Bailey, S. RNA-activated DNA cleavage by the Type III-B CRISPR–Cas effector complex. *Genes Dev.* **30**, 460–470 (2016).

37. Elmore, J. R. *et al.* Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes Dev.* **30**, 447–459 (2016).

38. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).

39. Özcan, A. *et al.* Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nat. Microbiol.* **4**, 89–96 (2019).

40. Tang, Y. & Fu, Y. Class 2 CRISPR/Cas: An expanding biotechnology toolbox for and beyond genome editing 06 Biological Sciences 0604 Genetics. *Cell Biosci.* **8**, 1–13 (2018).

41. Crawley, A. B., Henriksen, J. R. & Barrangou, R. CRISPRdisco: An Automated Pipeline for the Discovery and Analysis of CRISPR-Cas Systems. *Cris. J.* **1**, crispr.2017.0022 (2018).

42. Jiang, F. & Doudna, J. A. CRISPR – Cas9 Structures and Mechanisms. *Annu.Rev.Biophys* **46**, 505–529 (2017).

43. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 6091–6105 (2014).

44. Yan, W. X. *et al.* Functionally diverse type V CRISPR-Cas systems. *Science.* **363**, 88–91 (2019).

45. Stella, S. *et al.* Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* 1856–1871 (2018). doi:10.1016/j.cell.2018.10.045

46. Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science (80-.).* **360**, 436–439 (2018).

47. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell* **60**, 385–397 (2015).

48. Zhang, C. *et al.* Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d. *Cell* **175**, 212–223.e17 (2018).

49. Bondy-Denomy, J. & Davidson, A. R. When a virus is not a parasite: The beneficial effects of prophages on bacterial fitness. *J. Microbiol.* **52**, 235–242 (2014).

50. Goldberg, G. W., Jiang, W., Bikard, D. & Marraffini, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**, 633–637 (2014).

51. Sefcikova, J., Roth, M., Yu, G. & Li, H. Cas6 processes tight and relaxed repeat RNA via multiple mechanisms: A hypothesis. *BioEssays* **39**, 1–6 (2017).

52. Hatoum-Aslan, A., Maniv, I. & Marraffini, L. A. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc. Natl. Acad. Sci.* **108**, 21218–21222 (2011).

53. You, L. *et al.* Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. *Cell* **176**, 239–253 (2019).

54. Staals, R. H. J. *et al.* Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. *Mol. Cell* **52**, 135–145 (2013).

55. Hale, C. R. *et al.* RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139**, 945–956 (2009).

56. Zhang, J. *et al.* Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity. *Mol. Cell* **45**, 303–313 (2012).

57. Jia, N. *et al.* Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity. *Mol. Cell* **73**, 264–277 (2019).

58. Marraffini, L. A. & Sontheimer, E. J. CRISPR Interference Limits Horizontal Targeting DNA. *Science (80-.).* **322**, 1843–1845 (2008).

59. Han, W. *et al.* A type III-B CRISPR-Cas effector complex mediating massive target DNA destruction. *Nucleic Acids Res.* **45**, 1983–1993 (2017).

60. Liu, T. Y., Iavarone, A. T. & Doudna, J. A. RNA and DNA targeting by a reconstituted Thermus thermophiles Type III-A CRISPR-Cas system. *PLoS One* **12**, 1–20 (2017).

61. Zhang, J., Graham, S., Tello, A., Iu, H. & White, M. F. Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Res.* **44**, 1789–1799 (2016).

62. Guillerez, J., Lopez, P. J., Proux, F., Launay, H. & Dreyfus, M. A mutation in T7 RNA polymerase that facilitates. *PNAS* **102**, 5958–5963 (2005).

63. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci.* **108**, 10098–10103 (2011).

64. Anderson, E. M. *et al.* Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity. *J. Biotechnol.* **211**, 56–65 (2015).

CURRICULUM VITAE

Kaitlin Johnson, PhD

kjohn176@jhu.edu ◆ (260) 224-5896
615 North Wolfe Street, Baltimore, MD 21205

EDUCATION

Johns Hopkins University Bloomberg School of Public Health — Baltimore, MD
Ph.D., Biochemistry and Molecular Biology — March 2019

Manchester College (now Manchester University) — North Manchester, IN
B.A. *cum laude* Chemistry, Spanish — May 2012

RESEARCH EXPERIENCE

Johns Hopkins University, School of Public Health: Department of Biochemistry & Molecular Biology
Graduate student with Dr. Scott Bailey — 2012-present

- Determined the mechanism of host transcript identification in a Type III CRISPR-Cas system
- Assessed tolerance for mismatches between guide and target RNAs in a Type III CRISPR-Cas system
- Collaborated with Dr. Fengyi Wan to purify and characterize the secreted *E. coli* protease NleC
- Collaborated with Dr. Jiou Wang to purify a protein implicated in ALS, C9orf72

Manchester College: Department of Chemistry
Undergraduate research with Dr. Jeffrey Osborne — August 2011- May 2012

- Worked toward the synthesis, isolation, and analysis of styrene trimer

PUBLICATIONS

- **Johnson, K.**, Learn, B., Estrella, M. A., and Bailey, S. Target sequence requirements of a type III CRISPR-Cas immune system. *JBC* in press (2019).
- Liu, Y., Wang, T., Ji, Y. J., Johnson, K., Liu, L., **Johnson, K.**, Bailey, S., Suk, Y., Lu., Y., Liu, M., and Wang, J. A C9orf72-CARM1 axis regulates lipid metabolism under glucose starvation-induced nutrient stress. *Genes Dev*. **32**: 1380 – 1397 (2018).
- **Johnson, K.** and Bailey, S. The case of the mysterious messenger. *Nature*. **548**: 527–528 (2017).
- Hodgson, A., Wier, E. M., Fu, K., Sun, X., Yu, H., Zheng, W., Sham, H. P., **Johnson, K.**, Bailey, S., Vallance, B. A., and Wan, F. Metalloprotease NleC suppresses host NF-κB/inflammatory responses by cleaving p65 and interfering with the p65/RPS3 interaction. PLoS Pathog. **11**(3): 1-23 (2015).

SELECTED ORAL PRESENTATIONS

Johns Hopkins University Molecular Mechanisms League — May 2018
Host transcript recognition in a Type III-B CRISPR-Cas system — *Baltimore, MD*

Biochemistry and Molecular Biology Annual Departmental Retreat — April 2017
DNA cleavage activation in *Thermotoga maritima* — *Gettysburg, PA*
**Received award: Best Oral Presentation (2nd Place)*

Biochemistry and Molecular Biology Departmental Colloquium: Bailey Lab — October 2017
Host transcript recognition in *Thermotoga maritima* Cmr — *Baltimore, MD*

Biochemistry and Molecular Biology Annual Departmental Retreat — April 2016
RNA-activated DNA degradation by *Thermotoga maritima* Cmr — *Towson, MD*

POSTER PRESENTATIONS

Biochemistry and Molecular Biology Annual Departmental Retreat April 2018
Preventing autoimmunity in transcription-dependent CRISPR-Cas systems *Baltimore, MD*
Johnson, K. Learn, B., and Bailey, S.

Experimental Biology Conference April 2018
Preventing autoimmunity in transcription-dependent CRISPR-Cas systems *San Diego, CA*
Johnson, K. Learn, B., and Bailey, S.

TEACHING AND LEADERSHIP EXPERIENCE

Graduate Teaching Assistant

- Concepts of Molecular BiologySeptember – October 2017
- Introduction to Molecular Biology September – October 2016

Graded exams and held open office hours, reviews, and private tutoring sessions for MHS and PhD students

Rotation student mentoring

- PhD student mentor October – December 2018
- PhD student co-mentor January – March 2018

Master's student mentoring

- ScM student co-mentor June 2016 – April 2017
- MHS student mentor October – December 2016
- MHS student mentor September – December 2015

Undergraduate student mentoring

- Biophysics Research for Baltimore Teens student mentor June – August 2017
- Biophysics Research for Baltimore Teens student mentor June – August 2015

PhD student representative September 2014 – September 2016

- Attended faculty meetings as elected representative
- Held town hall meetings and relayed PhD student concerns and requests to faculty
- Helped to create the BMB PhD Club, a monthly meeting dedicated to career development

Departmental events committee co-chair May 2017 – January 2018

- Planned and set up department events, including annual picnic, holiday party, monthly happy hours
- Designed and sold departmental apparel to fundraise for events

SKILLS AND TECHNIQUES

- Recombinant protein expression and purification
- Enzymatic characterization
- Protein crystallization
- Bacterial cell culture
- *In vitro* transcription
- Autoradiography
- EMSA
- Molecular cloning (traditional, LIC, Gibson assembly, site-directed mutagenesis)
- Gel electrophoresis

RELEVANT COURSEWORK

Johns Hopkins University, School of Public Health

Genome Integrity and Cancer	2013

Johns Hopkins University, School of Medicine

Fundamentals of Protein Crystallography	2014
Molecular Biology and Genomics	2012
Biochemical and Biophysical Principles	2012
Macromolecular Structure and Analysis	2012
Computational Biology and Bioinformatics	2012

ACTIVITIES

Departmental events committee September 2013 – January 2018

- Helped to plan and set up events, including annual picnic, holiday party, monthly happy hours
- Appointed co-chair: May 2017 – January 2018

PhD recruitment weekend aide February 2013 – 2017

- Transported recruits to hotel and restaurants
- Escorted recruits on tour of Baltimore, around campus, and to interviews

REFERENCES

Scott Bailey (thesis advisor)
Associate professor
Department of Biochemistry and Molecular Biology
Johns Hopkins University Bloomberg School of Public Health
615 N. Wolfe Street, Baltimore, MD 21205
443-756-6791
scott.bailey@jhu.edu

Jennifer Kavran (thesis committee member)
Assistant professor
Department of Biochemistry and Molecular Biology
Johns Hopkins University Bloomberg School of Public Health
615 N. Wolfe Street, Baltimore, MD 21205
410-955-3671
jkavran@jhu.edu

Jungsan Sohn (thesis committee member)
Associate professor
Department of Biophysics and Biophysical Chemistry
Johns Hopkins University School of Medicine
725 N. Wolfe Street, Baltimore, MD 21205
443-287-1957
jsohn@jhmi.edu